

# **A FRAMEWORK AND EVALUATION OF CONVERSATION AGENTS**

**Ong Sing Goh**

B.A. Edu (Hons), University Science Malaysia

MSc, University of Manchester Institute of Science and Technology (UMIST)

**This thesis is presented for the degree of Doctor of Philosophy  
School of Information Technology  
Murdoch University  
2008**

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

---

Ong Sing Goh

## ABSTRACT

This project details the development of a novel and practical framework for the development of conversation agents (CAs), or conversation robots. CAs, are software programs which can be used to provide a natural interface between human and computers. In this study, ‘conversation’ refers to real-time dialogue exchange between human and machine which may range from web chatting to “on-the-go” conversation through mobile devices. In essence, the project proposes a “smart and effective” communication technology where an autonomous agent is able to carry out simulated human conversation via multiple channels. The CA developed in this project is termed “Artificial Intelligence Natural-language Identity” (AINI) and AINI is used to illustrate the implementation and testing carried out in this project. Up to now, most CAs have been developed with a short term objective to serve as tools to convince users that they are talking with real humans as in the case of the Turing Test. The traditional designs have mainly relied on *ad-hoc* approach and hand-crafted domain knowledge. Such approaches make it difficult for a fully integrated system to be developed and modified for other domain applications and tasks. The proposed framework in this thesis addresses such limitations. Overcoming the weaknesses of previous systems have been the key challenges in this study. The research in this study has provided a better understanding of the system requirements and the development of a systematic approach for the construction of intelligent CAs based on agent architecture using a modular *N*-tiered approach. This study demonstrates an effective implementation and exploration of the new paradigm of Computer Mediated Conversation (CMC) through CAs. The most significant aspect of the proposed framework is its ability to re-use and encapsulate expertise such as domain knowledge, natural language query and human-computer interface through plug-in components. As a result, the developer does not need to change the framework implementation for different applications. This proposed system provides interoperability among heterogeneous systems and it has the flexibility to

be adapted for other languages, interface designs and domain applications. A modular design of knowledge representation facilitates the creation of the CA knowledge bases. This enables easier integration of open-domain and domain-specific knowledge with the ability to provide answers for broader queries. In order to build the knowledge base for the CAs, this study has also proposed a mechanism to gather information from commonsense collaborative knowledge and online web documents. The proposed Automated Knowledge Extraction Agent (AKEA) has been used for the extraction of unstructured knowledge from the Web. On the other hand, it is also realised that it is important to establish the trustworthiness of the sources of information. This thesis introduces a Web Knowledge Trust Model (WKTM) to establish the trustworthiness of the sources.

In order to assess the proposed framework, relevant tools and application modules have been developed and an evaluation of their effectiveness has been carried out to validate the performance and accuracy of the system. Both laboratory and public experiments with online users in real-time have been carried out. The results have shown that the proposed system is effective. In addition, it has been demonstrated that the CA could be implemented on the Web, mobile services and Instant Messaging (IM). In the real-time human-machine conversation experiment, it was shown that AINI is able to carry out conversations with human users by providing spontaneous interaction in an unconstrained setting. The study observed that AINI and humans share common properties in linguistic features and paralinguistic cues. These human-computer interactions have been analysed and contributed to the understanding of how the users interact with CAs. Such knowledge is also useful for the development of conversation systems utilising the commonalities found in these interactions. While AINI is found having difficulties in responding to some forms of paralinguistic cues, this could lead to research directions for further work to improve the CA performance in the future.



# TABLE OF CONTENTS

<b>Abstract</b>	<b>iii</b>
<b>Table of Contents</b>	<b>v</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>List of Publications</b>	<b>xii</b>
<b>Contribution of the Thesis</b>	<b>xvii</b>
<b>Index of Figures</b>	<b>xx</b>
<b>Index of Tables</b>	<b>xxii</b>
<b>Index of Abbreviations and Acronyms</b>	<b>xxiii</b>
<b>CHAPTER 1: INTRODUCTION</b>	<b>1</b>
1.1 Overview	1
1.2 Ethical Considerations	6
1.3 Delimitations of the Thesis	7
1.4 Organisation of the Thesis	8
<b>CHAPTER 2: BACKGROUND</b>	<b>11</b>
2.1 Introduction	11
2.2 Artificial Intelligence (AI)	12
2.3 Turing Test and Loebner Prize	13
2.4 State-of-the-arts Conversation Agents System	17
2.4.1 Classical Conversation Agents System	21
2.4.1.1 ELIZA	21
2.4.1.2 PARRY	23
2.4.2 Conversation Agents in the Loebner Prize	25
2.4.2.1 ALICE	26
2.4.2.2 Jaberwacky	33
2.4.3 Commercial Conversation Agents System	35
2.4.3.1 Anna	36
2.4.3.2 Spleak	37
2.4.4 Tricks or AI	38
2.5 New Challenges	41
2.5.1 Natural Language Understanding	41
2.5.2 World Knowledge	42
2.5.3 Human-machine Interface	43
2.6 Summary	43

<b>CHAPTER 3: CONVERSATION AGENTS FRAMEWORK DESIGN</b>	<b>46</b>
3.1 Introduction	46
3.2 Conversation Agents Framework	46
3.2.1 Reusability	47
3.2.2 Modularity	48
3.2.3 Extensibility	48
3.2.4 Scalability	48
3.3 Conversation Agents' Features	49
3.3.1 Modules Integration	49
3.3.2 Domain Independent	49
3.3.3 Cross-Platform	50
3.4 <i>N</i> -tiered Architecture	51
3.5 AINI's Conversation Agents Architecture	52
3.5.1 AINI's Modified <i>N</i> -tiered Architecture	53
3.5.1.1 Channel Service Tier	55
3.5.1.2 Domain Service Tier	55
3.5.2 Agent Body ( <i>Client Tier</i> )	57
3.5.2.1 WebChat	58
3.5.2.2 MobileChat	60
3.5.2.3 MSNChat	62
3.5.2.4 Proxy Conversation Example 1	66
3.5.3 Agent Knowledge ( <i>Data Server Tier</i> )	69
3.5.3.1 Domain Knowledge Matrix Model (DKMM)	70
3.5.3.2 Open-Domain Knowledge Bases	72
3.5.3.3 Domain-Specific Knowledge Bases	76
3.5.3.4 Stimulus-Response Categories in AINI's Knowledge bases	76
3.5.3.5 Proxy Conversation Example 2	77
3.5.4 Agent Brain ( <i>Application Server Tier</i> )	79
3.5.4.1 Multilevel Natural Language Query	80
3.5.4.2 Spell Checker	81
3.5.4.3 Natural Language Understanding and Reasoning (NLUR)	83
3.5.4.4 Frequently Asked Questions Chat (FAQChat)	86
3.5.4.5 Index Search	87
3.5.4.6 Pattern Matching & Case-based Reasoning (PMCBR)	87
3.5.4.7 Supervised Learning Approach by Domain Expert	90
3.5.4.8 Proxy Conversation Example 3	92
3.6 Adaptability of the AINI's Framework into other Domain Application	95
3.7 AINI Compares with others Conversation Agents	98
3.7.1 Multilevel Natural Language Query	98
3.7.2 Spelling Correction	98
3.7.3 Implementation	99
3.7.4 Supervise Learning	101
3.7.5 Dynamic Responses	102
3.8 Summary	102
 <b>CHAPTER 4 AN ASSESSMENT OF THE TRUSTWORTHINESS OF KNOWLEDGE BASES FOR CONVERSATION AGENTS</b>	 <b>104</b>
4.1 Introduction	104
4.2 Trust and Methodology	106
4.3 Website as the Object of Trust	112

4.4	Trust Model	112
4.5	Web Knowledge Trust Model (WKTm)	113
4.5.1	Selecting Domain-Specific Web Knowledge	115
4.5.2	Seeding	116
4.5.3	Building a Corpus	118
4.5.4	Evaluating a Corpus	120
4.5.4.1	Log Likelihood Ratio	121
4.5.4.2	PageRank	123
4.5.4.3	Web Credibility	126
4.5.5	Trustworthiness Websites	130
4.6	Automated Knowledge Extraction Agent (AKEA)	132
4.6.1	Crawler	132
4.6.2	Wrapper	133
4.6.3	Text Categoriser	134
4.6.4	Syntactic Preprocessor	134
4.6.5	Semantic Parser	134
4.6.6	Semantic Interpreter	135
4.6.7	Query Engine	135
4.7	Summary	137
<b>CHAPTER 5: AN EVALUATION OF THE CONVERSATION AGENTS FRAMEWORK</b>		<b>139</b>
5.1	Introduction	139
5.2	An Evaluation of the Parsers	139
5.2.1	Performance of the Parsers	141
5.2.2	Accuracy of the Parsers	141
5.3	An Evaluation of the Performance Conversation Agents	145
5.4	A Comparison Response Quality for Query Systems: <i>Search Engine, Question-answering and Conversation System</i>	149
5.5	Summary	153
<b>CHAPTER 6: AN ANALYSIS OF THE LINGUISTIC FEATURES FROM REAL-TIME HUMAN-MACHINE INTERACTION</b>		<b>155</b>
6.1	Introduction	155
6.2	Experimental Setting	157
6.2.1	Participants and Corpus	158
6.2.2	Measures	160
6.2.3	Conversation Logs	163
6.3	Domain Knowledge and Conversation Topics	166
6.4	Linguistic Analysis	169
6.4.1	Word Frequency Analysis	169
6.4.2	Lexical Analysis	171
6.4.2.1	Humanness of Conversation with Pronouns	171
6.4.2.2	Contracted Words	172
6.4.3	Text Complexity	174
6.5	Summary	178

<b>CHAPTER 7: AN ANALYSIS OF THE PARALINGUISTIC CUES FROM</b>	<b>181</b>
<b>REAL-TIME HUMAN-MACHINE INTERACTION</b>	
7.1 Introduction	181
7.2 Paralinguistic Analysis	183
7.2.1 Intonations with Interjections and Fillers	184
7.2.2 Abbreviations with Acronyms and Shorthand	187
7.2.3 Facial Expressions with Emoticons or Smiley	190
7.2.4 ASCII Arts	193
7.2.5 Prosody Markers with Pauses and Voice Pitch	194
7.3 Summary	196
 <b>CHAPTER 8: CONCLUSIONS</b>	 <b>197</b>
8.1 Introduction	197
8.1.1 Summary	197
8.1.2 Contributions	199
8.2 Limitations	201
8.2.1 Agent Knowledge Issues	201
8.2.2 Agent Brain Issues	202
8.2.3 Agent Body Issues	203
8.3 Future Work and Directions	203
 <b>REFERENCES</b>	 <b>205</b>
 <b>APPENDIXES</b>	 <b>224</b>
<b>APPENDIX A: Online Consent Form for Real-time Experiment Approved by Human Research Ethics Committee (HREC), Murdoch University</b>	<b>224</b>
 <b>APPENDIX B: An Example of the Full Sentence Parsing using Natural Language Understanding and Reasoning (NLUR) Module for a sentence “<i>Bird flu did occur in which countries?</i>”</b>	 <b>230</b>
 <b>APPENDIX C: Google PageRank™ Checksum Calculator - Assigning Numerical Weightings to Hyperlinked Documents Indexed by a Google Search Engine</b>	 <b>236</b>

## **ACKNOWLEDGEMENTS**

First of all I would like to take this opportunity to express my most heartfelt gratitude to my supervisor Associate Professor Lance Chun Che Fung for his support, guidance, comments and advice throughout the course of my study. From day one, Professor Fung always put the development of generic skills such as intellectual understanding and judgment, problem solving skills, critical thinking skills, personal and interpersonal skills, communication skills, cooperative and collaborative teamwork, and leadership as the highest priority during his supervision. He always stresses the fact that the thesis is not the only purpose of the study, but it is a process that the skills I learnt throughout the study is more important as they will help me in every aspects of my life. He is not only my mentor, but also a friend. I owe Professor Fung a lot in my development as a researcher and an academic.

I am also greatly indebted to Associate Professor Kevin Wong and Professor Arnold Depickere. They have always offered me a stimulating, supportive and comfortable environment to grow. I am grateful for their guidance and encouraging outside-the-box thinking and high standards which reflect their genuine support and supervision.

Besides my supervisors, there are many other scholars behind the writing of this thesis worthy of my mention. First, for the idea of this research, I must say that I owe a lot of insight to Professor Lotfi A. Zadeh (University of California, Berkeley), the Father of Fuzzy Logic whom I met during my paper presentation at the International Conference on Intelligent Agents, Web Technology and Internet Commerce (IAWTIC2003) Vienna, Austria. He taught me a great deal about “computing with words and perceptions” and provided important pieces of advice that have guided me into this research. Second, for the knowledge of computational linguistics, I owe a debt of thanks to Professor Allan Ramsay, my former supervisor who gave me the inspiration to start work on formal linguistics and

Natural Language Processing during my postgraduate study in the United Kingdom at the University of Manchester Institute of Science and Technology (UMIST), which has led to this research. Third, many thanks to Professor Srisakdi Charmonman (Chairman of the Board and CEO at the College of Internet Distance Education, Assumption University) for inviting me to give a Keynote Address. I appreciated the chance to discuss my ideas at the 2<sup>nd</sup> International Conference on eLearning for Knowledge-Based Society (ELAP2005), Bangkok.

I also wish to express my gratitude to the developers of Conversation Agents (CAs) or ‘chatterbots’ around the world at Robitron,<sup>1</sup> including winners of the Loebner Prize, Dr. Richard S. Wallace, Robby Garner, Kevin Copple and Rollo Carpenter to name a few for their help in the experiments, encouragement and fruitful discussions throughout my academic journey.

This thesis would not have been possible without the commitment of some special people.

I would like to thank all of them for their help:

- Dr. Marco Baroni for providing the BootCAT toolkit, a simple utility to bootstrap corpora and terms from the Web.
- Dr. Hugh Gene Loebner for providing the transcripts from the Loebner Prize competition for use in this project. The Loebner Prize is the premier international competition on Artificial Intelligence concerning conversation agents.
- Rion Snow from Stanford AI Lab for providing the MINIPAR visualisation tool.
- Dr. Richard S. Wallace, for providing the Annotated ALICE AIML (AAA) stimulus-categories knowledge base, and awarding life-time ALICE Foundation membership to access ALICE Silver Edition for this study.
- Professor Ted Pedersen for providing Ngram Statistics Package (NSP), a computational linguistic tool to identify word *n*-grams that appear in large corpora with minimal effort.
- Adi Sideman, President and CEO of Oddcast<sup>TM</sup>, who granted the VHost Studio authoring tool that allows us to create and embed customised animated characters

---

<sup>1</sup> <http://groups.yahoo.com/group/robitron/>

within HTML pages and speak any text in real time, with accurate lip synchronisation to deliver truly dynamic conversational character experience.

My research would not have become a reality without the financial assistance from University Technical Malaysia Melaka (UTeM) and a Murdoch International Scholarship, a three-year stipend that enabled me to give my full attention to my research. I would not have come this far without the opportunity and financial support from them.

In addition to the above, others too have helped me in one way or another. I thank Professor Shahrin Sahib (Dean, School of Information Technology and Communication, Technical University of Malaysia Melaka) who trusted me to serve the faculty and indirectly encouraged me to continue research in intelligence agents. This research will not be possible without the help of Mr Wilson Wong, who contributed as a co-author for numerous papers published in proceedings and journals.

No words are sufficient to express my true appreciation to my lovely wife Bee Poo Tee and my three sons Vincent Yee Hang Goh, Yee Shien Goh and Yee Jing Goh for their blessings, gratitude and tireless support. Not forgetting, my mum Lee Chu Niyu back in Melaka, Malaysia, for wishing me luck and success while waiting for my return.

Last but not the least, big thanks to Miriam Everall and anonymous referees for the very careful readings of many manuscripts, helpful suggestions and help in getting them published. I also would like to thank all the staff and postgraduate students in the School of Information Technology for their friendship and assistance during my research study. Thanks are also due to all subjects that attended my experiments. Without their participation, I would not have been able to advance my research.

## LIST OF PUBLICATIONS

The following papers have reported the progress and results of work related to this thesis. Most of the earlier work was focused on question-answering systems and knowledge extraction. Since the CA framework incorporates an extendable design, subsequently, the focus of the work was shifted to its adaptation to other applications domain and evaluation of the conversation system. There are a total of 31 publications (three submitted for review) which include two book chapters, twelve journal articles and seventeen papers in proceedings of international conferences<sup>2</sup>.

### Book Chapters

- B1. O. S. Goh, C. C. Fung, C. Ardil, K. W. Wong, and A. Depickere, "An Analysis of Man-machine Interaction in Instant Messenger", in *Advances in Communication Systems and Electrical Engineering*, Huang, Xu; Chen, Yuh-Shyan; Ao, Sio-Iong (Eds.), Vol. 4 pp. 197-210, 2008, Springer United State, ISBN: 978-0-387-74937-2.
- B2. O. S. Goh and C. C. Fung, "Automated Knowledge Extraction from Internet for a Crisis Communication Portal" in *Fuzzy Systems and Knowledge Discovery*, Lipo Wang, Yaochu Jin (Eds.), Vol. 3614, pp. 1226-1235, 2005, Springer Berlin, ISBN: 978-3-540-28331-7

### Journal Papers

- J1. O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "Linguistic Analysis of an Instant Messaging Conversation Agent Corpus," submitted for review to the *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, IEEE Press, 2008, ISSN: 1083-4427.

---

<sup>2</sup> Lists of publications related to this thesis can be found at <http://osgoh.ainibot.org>



- J2. O. S. Goh, C. C. Fung, "Building an Intelligent Conversation Agent's Domain Knowledge based on a Web Knowledge Trust Model (WKTM)", submitted for review to the Special Issue on Knowledge Discovery for Web Intelligence in *ACM Transactions on Knowledge Discovery from Data*, ACM Press, 2008, ISSN: 1556-4681
- J3. O. S. Goh, C. C. Fung, and K. W. Wong, "Paralinguistic Cues in IM conversation Robots," submitted for review in the *International Journal of Social Robotics*, Springer Netherlands, 2008, ISSN: 1875-4791.
- J4. O. S. Goh, C. C. Fung, and A. Depickere, "Domain Knowledge Query Conversation Bots in Instant Messaging (IM)", *Knowledge-Based Systems*, Vol. 21, Issue 7, pp. 681-691, October 2008, Elsevier, ISSN: 0950-7051
- J5. O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "Query Based Intelligent Web Interaction with Real World Knowledge," *Special Feature: Intelligent Web Interaction, New Generation Computing*, S. Yamada and T. Murata (Eds.), Vol. 26, No 1, pp. 3-22, 2008, Ohmsha, Ltd. and Springer, ISSN: 0288-3635.
- J6. O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, A., Embodied Conversational Agents for H5N1 Pandemic Crisis", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.11, No.3, pp. 282-288, 2007, Fuji Press, ISSN: 1343-0130.
- J7. O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "Multilevel Natural Language Query Approach for Conversational Agent System", *IAENG International Journal of Computer Science*, Vol. 33, No 1, pg. 7-13, 2007, International Association of Engineers, ISSN :1819-656X.
- J8. O. S. Goh, C. C. Fung, A. Cemal K. W. Wong, A. Depickere, "A Crisis Communication Network Based on Embodied Conversational Agents System with Mobile Services", *International Journal of Information Technology*, Vol. 3, No 1, pp. 257-266, 2006, World Academy of Science, Engineering and Technology, ISSN :1305-2403.
- J9. O. S. Goh, A. Cemal, W. Wong, C. C. Fung, "A Black-box Approach for Response Quality Evaluation Conversational Agent System", *International Journal of Computational Intelligence*, Vol. 3, No 3. pp. 195-203, 2006, World Academy of Science, Engineering and Technology, ISSN :1304-2386

- J10. O. S. Goh, C. C. Fung, A. Depickere, and H. S. Lau, "Object-Based Learning with Concept Mapping Embodied Intelligent Agent", *International Journal of the Computer, The Internet and Management*, Special Issue, Vol. 12, No 1. pp. 30.1 - 30.7, August 2005, ISSN 0858-7027
- J11. O. S. Goh, A. Cemal, W. Wong, S. Sahib, "Response Quality Evaluation in Heterogeneous Question Answering System: A Black-box Approach", *Transactions on Engineering, Computing and Technology*, Vol. 9, pp. 49-54, November 2005, World Academy of Science, Engineering and Technology, ISSN 1305-531
- J12. O. S. Goh, C. C. Fung, and L. M. Ph'ng, "Intelligent Agents for an Internet-based Global Crisis Communication System", *Journal of Technology Management And Entrepreneurship*, Vol. 2, No 1, pp. 67-78, July 2005, ITME Press, ISSN: 1675-8404

### **Conference Proceedings**

- P1. O. S. Goh, C. C. Fung, and K. W. Wong, "VisualChat: A Visualisation Tool for Human-Machine Interaction", in *The 2008 IEEE/WIC/ACM Intelligent Web Interaction Workshops (IWI'08)*, IEEE Computer Society Press, 9 – 12 December, Sydney, Australia, Submitted for review.
- P2. O. S. Goh, C. C. Fung, "AINI - Embodied Conversation Agent Applicable for Interactive Games", in *The 7th WSEAS International Conference on Applied Computer and Applied Computational Science (ACACOS '08)*, pp. 272 – 277, WSEAS Press, 6-8 April 2008, Hangzhou, China
- P3. O. S. Goh, C. C. Fung, "Acquiring Trustworthy Knowledge for Conversation Agents based on a Web Knowledge Trust Model", in *The IAENG International Conference on Internet Computing and Web Services*, International Association of Engineers, 18-20 March 2008, Hong Kong
- P4. O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "Using Gunning-Fog Index to Assess Instant Messages Readability from ECAs", in *The 3rd International Conference on Natural Computation (ICNC'07)*, Vol. 5, pp 480-486, IEEE Computer Society Press, 24-27 August 2007, Haikou, China

- P5. O. S. Goh, C. C. Fung, A. Depickere, and K. W. Wong, "An Analysis of Corpus from Human Computer Exchanges using MSN Messenger", in *The IAENG International Conference on Internet Computing and Web Services*, International Association of Engineers, 21-24 March 2007, Hong Kong
- P6. O. S. Goh, A. Depickere, C. C. Fung, and K. W. Wong, "Domain Matrix Knowledge Model for Embodied Conversation Agents", in *The 5th International Conference on Research, Innovation & Vision for the Future (RIVF'07)*, 5-9 March 2007, Hanoi, Vietnam
- P7. O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "An Embodied Conversational Agent for Intelligent Web Interaction on Pandemic Crisis Communication", in *The IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IATW'06)*, IEEE Computer Society Press, pp. 397-400, 18-22 December, 2006, Hong Kong
- P8. O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "Towards A More Natural and Intelligent Interface with Embodied Conversation Agent", in *The 2006 international conference on Game Research and development*, ACM International Conference Proceeding Series, ACM Press, Vol. 223, pp. 177 – 183, Perth, Western Australia, 4-6 December 2006
- P9. O. S. Goh, K. W. Wong, A. Depickere, and C. C. Fung. "Empowering ECAs in Handheld Devices", in *The Joint 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on Advanced Intelligent Systems (SCIS & ISIS 2006)*, 20 - 24 September, Tokyo, Japan
- P10. O. S. Goh, A. Depickere, C. C. Fung, and K. W. Wong, "Top-down Natural Language Query Approach for Embodied Conversational Agent", in *The International MultiConference of Engineers and Computer Scientists 2006*, pp. 470-475, 27-29 June 2006, Hong Kong
- P11. O. S. Goh, C. C. Fung, A. Depickere, K. W. Wong, and W. Wong, "Domain Knowledge Model for Embodied Conversation Agent", in *The 3rd International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS 2005)*, December 2005, Singapore
- P12. O. S. Goh, C. C. Fung, A. Depickere, W. Wong, S. Sahib, "Intelligent Question Answering with Natural Language Understanding and Network-based Advanced

- Reasoning”, in *The International Conference on Intelligent Technologies (InTech'05)*, pp. 213 -222, December 2005, Phuket, Thailand
- P13. O. S. Goh, A. Cemal, W. Wong., S. Sahib, “Response Quality Evaluation in Heterogeneous Question Answering System: A Black-box Approach”, in *The 7th International Conference on Enformatika Systems Sciences and Engineering (ESSE 2005)*, November 2005, Istanbul, Turkey
- P14. O. S. Goh, C. C. Fung, A. Depickere, and H. S. Lau, “Embodied Intelligent Agent in eLearning”, in *The International Conference on eLearning for Knowledge-based Society*, 4-7 August 2005, Bangkok, Thailand
- P15. O. S. Goh, C. C. Fung, “Automated Knowledge Extraction from Internet for a Crisis Communication Portal”, in *The Second International Conference, FSKD 2005 (Part II)*, Springer Berlin, pp. 1226-1235, 27-29 August 2005, Changsha, China,
- P16. W. Wong, H. Basiron, S. Sahib, and O. S. Goh, “Intelligent Responses Through Network-Based Answer Discovery with Advanced Reasoning”, in *The Fourth IASTED International Conference on Computational Intelligence (CI'05)*, 4-6 July 2005, Calgary, Alberta, Canada.
- P17. W. Wong, S. Sahib, and O. S. Goh, "Evaluation of Response Quality for Heterogeneous Question Answering Systems", in *The IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, IEEE Computer Society Press, pp. 610-613, 19-22 September 2005, Compiègne, France.

## **CONTRIBUTIONS OF THE THESIS**

The contributions in this thesis which have been published and reported are described below and summarised in Table 1.1.

A survey and review of various techniques in the development of CA systems has been completed. The work has been published in paper P8. Conference paper P8 was later extended to journal paper J5, which has been described in Chapter 2. This is a paper on the state-of-the-art development of the discipline. The paper presents the results from the initial literature study on conversation systems and how evaluation has been conducted with respect to the “naturalness” and “humanness” of the human-machine conversation as required in Turing Test (TT).

The development of the new CA framework design forms a part of Chapter 3. The work has been reported in papers P10, P13 and P14. These three conference papers have been extended to journal papers J7, J10 and J11 respectively. Paper J10 was a keynote address presented at the International Conference on eLearning for Knowledge-based Society 2005. In addition, papers P5 and P10 have also received the Best Paper Awards at the International Conference on Internet Computing and Web Services in 2007, and the International MultiConference of Engineers and Computer Scientists in 2006 respectively. Papers J6 and J12 described the contribution of the applicability and adaptability of the AINI’s framework in terms of specific domains relevant to the SARS epidemic and bird flu pandemic.

During the writing of papers P6, P7 and P11, it became obvious that the publicly available Google API (Application Programming Interface) and Google PageRank have great potential in identifying unbiased seeds and corpora for building the CAs’ knowledge bases.

Papers J4 and J5 described the experiments with Google API and Google PageRank as the main sources from which trustworthy CAs' knowledge bases were established. Paper journal J5 was published in the special issue on *Intelligent Web Interaction*, extended from P7 and it showed that Google API can be used to simplify the information discovery process. This paper proposed the Web Knowledge Trust Model (WKTm) to determine the trustworthiness of relevant sources from the Web. Paper P15 revealed a novel approach, the Automated Knowledge base Extraction Agent (AKEA), and this constitutes the core contribution described in Chapter 4. This paper was also extended to book chapter B2.

The contribution in Chapter 5 is the establishment of a baseline for evaluating CAs in comparison to other query systems such as search engines, question-answering systems and conversation systems. The comparison was based on qualitative and quantitative approaches, and it also gave an insight into the performance of the natural language parsers. Paper P13 was a report from evaluating the quality of the query systems. This approach can be used as a benchmark for evaluating new systems in other domains. Paper P13 was subsequently extended to journal paper J9.

Chapter 6 and 7 complete the research work with an evaluation of the real-time human-machine interaction and the findings have been reported in papers P1 to P5, J1 to J4 and B1. These papers described the rationale and the results of public real-time experiment evaluation based on unconstrained domain and unrestricted duration. The empirical approach was based on the analysis of a number of conversation logs collected from human-machine interaction via MSN Messenger. The analyses include an extensive account of observed dialogue phenomena, which include linguistic features and paralinguistic cues of the human-machine utterances, as well as the topics of interest.

**Table 1.1: Summary of the Contribution of the Thesis**

CHAPTER	CONTRIBUTIONS	PAPER NO
<b>Background</b>	Literature survey on previous research work from classical CAs, Loebner Prize CAs to commercial CAs.	P8
<b>Conversation Agents Framework Design</b>	Proposal of a modified <i>N</i> -tiered architecture that provides reusable, extensible, scalable, and modular (RESM) design for heterogeneous CAs framework.	P14, J12
	<b>AGENT BRAIN (<i>Application Server Tier</i>)</b> The development of a novel top-down multi-level natural language query approach.	P10, P12, P16, J17
	<b>AGENT KNOWLEDGE (<i>Data Server Tier</i>)</b> This thesis introduces a Web Knowledge Trust Model (WKTm) to establish Conversational Agents knowledge which consists of Open-domain and Domain-specific knowledge base. The main contribution of this model is the proposal and development of a Domain-specific knowledge from trustworthiness online documents using Automated Knowledge Extraction Agent (AKEA).	P6, P7, P11
	<b>AGENT BODY (<i>Client Tier</i>)</b> Proposal and development of a multiple-channel communication approach for greater CAs autonomy.	P9, J6, J8, J10
<b>An Assessment of the Trustworthiness of Knowledge Bases for Conversation Agents</b>	Through Google API, Google PageRank and Web Credibility, the World Wide Web is used as the main resource to find and extract trustworthy web pages using the proposed Web Knowledge Trust Model (WKTm). Automated Knowledge Extraction Agent (AKEA) is used to retrieve and dynamically construct trusted Web knowledge from semi-structured data.	P3, P15, J2, J4, J5, B2
<b>An Evaluation of the Conversation Agent Framework</b>	Short-term lab-based and controlled experiments are used to verify the proposed framework design. The evaluation demonstrated possible solutions to evaluate the quantitative performance and accuracy of the parsers; and response quality of the AINI conversation system.	P2, P13, P17, J9, J11
<b>An Analysis of the Linguistic Features from Real-time Human-Machine Interaction</b>	VisualChat tools have been developed to visualise the linguistic features and paralinguistic cues of the conversation between human and CAs in the real-time experiment. Results from the experiment showed that human and machines can communicate better in unrestricted domain, without a time limit and unconstraint setting.	P1, P4, P5, J1, B1
<b>An Analysis of the Paralinguistic Cues from Real-time Human-Machine Interaction</b>	The study also observed that human participants or AINI's buddies expressed their ideas and feeling through paralinguistic cues in the IM environments. By incorporating this feature, AINI is providing better and human-like conversations with the users.	J3

## INDEX OF FIGURES

Figure 1.1:	Conversation Agents Development Process	5
Figure 2.1:	Artificial Intelligence Categories	12
Figure 2.2:	Artificial Intelligence Timelines	13
Figure 2.3:	The Turing Test (TT)	14
Figure 2.4:	Conversation between ELIZA and Patient	22
Figure 2.5:	Decomposition and Reassembly Rules in ELIZA	22
Figure 2.6:	ELIZA Converse with PARRY	25
Figure 2.7:	ALICE Converse with ELIZA	28
Figure 2.8:	ALICE Chatting with Human	29
Figure 2.9:	Example Conversation between ALICE and Jabberwacky	34
Figure 2.10:	Example Conversation with Anna	37
Figure 2.11:	Example Conversation with Spleak	38
Figure 3.1:	Comparison Two-tiered vs Three-tiered Client-Server Architecture	52
Figure 3.2:	AINI's Modified <i>N</i> -tiered Architecture	54
Figure 3.3:	AINI's Conversation Agent Architecture	56
Figure 3.4:	Personalise User Interface	59
Figure 3.5:	Examples to illustrate AINI's Scalable Interface	60
Figure 3.6:	SMSChat Interface	61
Figure 3.7:	WAPChat Interface	61
Figure 3.8:	PDACHat Interface	61
Figure 3.9:	AINI and MSN Authentication Process	63
Figure 3.10:	MSNDesktopChat Interface	64
Figure 3.11:	MSNWebChat Interface	64
Figure 3.12:	MSNMobileChat Interface	65
Figure 3.13:	AINI's Communication Channel Module	66
Figure 3.14:	An example illustrating a user communicates with Live Agent (human) through MSN Messenger	67
Figure 3.15:	An example of Proxy Conversation from a Financial Portal based on WebChat	68
Figure 3.16:	Domain Knowledge Matrix Model (DKMM)	70
Figure 3.17:	An example of Proxy Conversation on SARS based on Domain Knowledge	78
Figure 3.18:	Multilevel Natural Language Query	81
Figure 3.19:	Natural Language Understanding and Reasoning Architecture	84
Figure 3.20:	AIML Categories and Pattern-matching	89
Figure 3.21:	AIML Symbolic Reduction (SRAI) Technique	90
Figure 3.22:	An Example of Proxy Conversation Log on H5N1 based on the NL-Query	94
Figure 3.23:	An Adaptive AINI Framework Architecture	96
Figure 3.24:	An Example of Migration from SARs to Bird Flu Domain Application	96
Figure 3.25:	Deploying AINI object into the Existing Website	97
Figure 3.26:	AINI Customisation Profiles	97
Figure 3.27:	Spell Check in AIML	99



Figure 3.28:	Random Responses Categories in AIML	100
Figure 4.1:	Web Knowledge Trust Model (WKTm)	114
Figure 4.2:	Comparing distribution of seed words between the smaller set data Pandemic Corpus with the Google Large-scale Corpus	120
Figure 4.3:	An Example of PageRank Corresponding to Web pages	124
Figure 4.4:	Top Ten PageRank Scale for the Bird Flu Domain	125
Figure 4.5:	Comparing Trustworthiness of Top 10 Websites related to the Bird Flu Domain	131
Figure 4.6:	AKEA Architecture	132
Figure 4.7:	Crawler's Configuration Interface	133
Figure 4.8:	Ontology Structure in AKEA	136
Figure 4.9:	Name Entity Recognition in AKEA	136
Figure 5.1:	Parse output of the CMU Link Grammar	142
Figure 5.2:	Dependency Graphs generated by CMU Link Grammar	142
Figure 5.3:	Parse output of the Stanford parser	143
Figure 5.4:	Dependency Graphs generated by Stanford Parser	143
Figure 5.5:	Parse output of the X-MINIPAR	143
Figure 5.6:	Dependency Graphs generated by X-MINIPAR	144
Figure 5.7:	Response time for AINI, ALICE, ALICE Silver edition and ELIZA	146
Figure 5.8:	Experimental Design Interface for Response Quality	150
Figure 6.1:	The Visualisation Pipeline	161
Figure 6.2:	An Example of Visualisation Chat between AINI and Human in IM using VisualChat	162
Figure 6.3:	A typical Single Session Conversation between AINI and user U1025	165
Figure 6.4:	Frequency of Conversation Topics	166
Figure 6.5:	Frequency of AINI's Responses based on Domain Knowledge Bases	167
Figure 6.6:	Visualisation of the Lexical Features used in the IM Conversation Agents	172
Figure 6.7:	A comparison of the Frequency of Contracted Verbs used in BNC and IM Conversation Agents	173
Figure 6.8:	Gunning-Fog Index from different Corpora	176
Figure 6.9:	A Sample of TRAINS Conversation Log	177
Figure 6.10:	A Sample of Swhack IRC Conversation Log	177
Figure 7.1:	A typical Single Session Conversation between AINI and user U1037	183
Figure 7.2:	Frequency list of Interjections and Fillers in IM Conversation Agents	185
Figure 7.3:	Visualisation of the Paralinguistic Cues Populated in IM Conversation Agents based on One-word Frequency	187
Figure 7.4:	Universal expression from Human to Computer Communication	191
Figure 7.5:	Frequency of Smiley and Emoticons used in the IM Conversation Agents	192
Figure 7.6:	ASCII Arts used in IM Conversation Agents	193
Figure 7.7:	Prosody Markers used in IM Conversation Agents	195

## INDEX OF TABLES

Table 1.1:	Summary of the Contribution of the Thesis	xix
Table 2.1:	Advances in Computer Technology	15
Table 2.2:	Advances in Conversation Agents in the Last Forty Years	18
Table 2.3:	A list of the Loebner Prize Winners from 1991 – 2007	26
Table 2.4:	Conversation Agents' Tricks	40
Table 3.1:	Number of Factoid Questions from TREC 8, 9 and 11	75
Table 3.2:	AINI's Stimulus-response Categories	77
Table 4.1:	Comparison of Trust Types and Measures in Previous Research	107
Table 4.2:	Website Trust Type and Methodology in Previous and Current Research	109
Table 4.3:	Comparing number of hit results from BNC and Google's Corpora using the set of unigram and bigram seed words	117
Table 4.4:	Contingency Table for Word Frequencies	122
Table 4.5:	Log-likelihood Ratios for Pandemic Corpus vs Google large-scale Corpus	123
Table 4.6:	Frequency of the Comment Topics for Website Credibility	126
Table 4.7:	An Example of Comments Related to Trustworthiness of the Website Information	127
Table 4.8:	Website Rankings based on Google's PageRank and Stanford's Web Credibility	129
Table 5.1:	Performance Test for X-MINIPAR, CMU Link Grammar and the Stanford Parser	141
Table 5.2:	Average response time and standard deviation for AINI, ALICE, ALICE Silver Edition and ELIZA	147
Table 5.3:	Natural Language Query Components in AINI compare to ELIZA and ALICE conversation systems	148
Table 5.4:	Responses from popular Search Engines – Google and Yahoo	151
Table 5.5:	Responses from popular Question-answering Systems	151
Table 5.6:	Responses from popular Conversation Systems	151
Table 6.1:	Frequency of Word from Conversation Logs	160
Table 6.2:	Top Ten Words used in Shakespeare, BNC and IM Conversation Agents	170
Table 6.3:	Frequency List of Pronouns used in BNC and IM Conversation Agents	171
Table 6.4:	A Comparison of the Frequencies of Contracted Verbs used in BNC and IM Conversation Agents	173
Table 6.5:	Gunning Fox Index with Unique word, Average word and Lexical Density from different Corpora	175
Table 7.1:	Log likelihood Ratio of Interjections and Fillers	185
Table 7.2:	Short forms used in IM Conversation Agents	188
Table 7.3:	Facial Expressions with Emoticons or Smileys used in IM Conversation Agents	191

## INDEX OF ABBREVIATIONS AND ACRONYMS

3G	Third Generation Protocol
AAA	Annotated ALICE AIML
AI	Artificial Intelligence
AIM	AOL Instant Messenger
AIML	Artificial Intelligence Markup Language
AINI	Artificial Intelligence Natural-language Identity
AKEA	Automated Knowledge Extraction Agent
ALICE	Artificial Linguistic Internet Computer Entity
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
BNC	British National Corpus
CA	Conversation Agent
CBR	Case Base Reasoning
CCNet	Crisis Communication Network
CMC	Computer-mediated Communication
CONV	Conversation
DKMM	Domain Knowledge Matrix Model
ECA	Embodied Conversation Agent
FAQ	Frequency Ask Question
GNU	General Public License
GPRS	General Packet Radio Service
HCI	Human-computer Interface
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol over Secure Sockets Layer
IE	Information Extraction
IG	Imitation Games
IM	Instant Messaging
IR	Information Retrieval
IRC	Internet Relay Chat
LAMP	UNIX, Apache, MySQL and Perl
LL	Log-likelihood
LIST	List Processing, a functional programming language
MIT	Massachusetts Institute of Technology
MMS	Multimedia Messages System

MSN	Microsoft Network
MSNP	Mobile Status Notification Protocol
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
NL-Query	Natural Language Query
NLU	Natural Language Understanding
NLUR	Natural Language Understanding and Reasoning
OS	Operating System
OSI	Open Source Initiative
PDA	Personal Digital Assistance
PERL	Practical Extraction and Reporting Language
PMCBR	Pattern Matching and Case Based Reasoning
POS	Part-of-Speech
QA	Question-answering
RESM	Reusable, Extensible, Scalable and Modular
SARS	Severe Acute Respiratory Syndrome
SMS	Short Messages System
SOA	Service-Oriented Architecture
TCP/IP	Transmission Control Protocol/Internet Protocol
TOS	Task Oriented Speech
TREC	Text Retrieval Conference
TT	Turing Test
URL	Uniform Resource Locator
US	United States
UTeM	University Technical Malaysia Melaka
WHO	World Health Organisation
WiFi	Wireless Fidelity
WKTM	Web Knowledge Trust Model
WWW	World Wide Web
XML	Extended Markup Language

# CHAPTER 1

## INTRODUCTION

### 1.1 Overview

Communication technologies have advanced greatly since the invention of the telegraph in 1794<sup>1</sup>. They have played important roles in connecting people and exchange of information with one another. In particular, fixed lines or mobile telephone systems have allowed users to carry out real-time conversation anytime and almost anywhere. However, advancements in computer and Internet technologies have taken the meaning of ‘conversation’ to a new dimension. Conversation is no longer limited to exchanges between human and human, but it can now describe communication between humans and computers. A software application that is capable of simulating human conversation is known as a conversation agent (CA). Providing CAs with knowledge and intelligence has allowed them to be used in several practical applications. For example, a CA is being advertised as a digital friend called Virtual Personalities by Verbot<sup>2</sup>. Other examples of applications for CAs are: therapists, counselors, teachers, trainers, salespersons or corporate representatives; such as those were being employed by companies are Spleak<sup>3</sup>, Extempo<sup>4</sup>, Artificial Life<sup>5</sup>, Kiwilogic<sup>6</sup>, and Oddcast Inc<sup>7</sup>.

Recently, there have been signs of great interest in the development of CAs driven by the needs of commercial applications. Many e-service providers are motivated to

---

<sup>1</sup> <http://inventors.about.com/od/tstartinventions/a/telegraph.htm>

<sup>2</sup> <http://www.verbots.com>

<sup>3</sup> <http://www.spleak.com>

<sup>4</sup> <http://www.extempo.com>

<sup>5</sup> <http://www.artificial-life.com>

<sup>6</sup> <http://www.kiwilogic.com>

<sup>7</sup> <http://www.oddcast.com>

incorporate natural language interfaces in their websites in order to improve customer relationships, resulting in an increase of sales. The agents used are human-like avatars acting as representatives for their respective companies. However, these agents are only focused on the interface to attract the viewers; rather than maintaining an ongoing interaction with the users.

Previous research works (discussed in Section 2.4), from classical CAs, through CAs in the Loebner Prize competition, to commercial CAs, have invented many subsets of conversation systems. These systems have all been developed for specific purposes, such as artificial intelligence (AI) research, to take part in the competition or for commercial use. For example, the development of Loebner Prize CAs could have been stimulated by the prize money of up to \$100,000 offered to the developers of the first system which can pass the Turing Test (TT) [5]. In general, this solution approach was custom-designed for the specific problem, and it is non-generalised. Such a solution may not be easily adapted to other purposes.

The primary aim of this study is to propose a framework for the development of reusable, extensible, scalable, and modular design CAs called AINI (Artificial Intelligent Natural-language Identity). The developed framework is based on agent architecture using a modular  $N$ -tiered approach. Under this architecture, different CAs handling different domains can be developed independently, and cooperate with one another to respond to the user's requests. Our CA's framework is primarily a client-server architecture that is capable of interacting with a user via text, speech and animated avatars to present information in a web browser environment, mobile service or computer-mediated communication (CMC) application. The proposed framework is not only limited by the communication channels stated above, it could also be extended

for use in future devices with an appropriate interface. The proposed new framework gives AINI-enabled applications the ability to understand and react to users based on user interactions, and to extract meaning from free-form text input.

This thesis addresses several challenges that arise in the development of practical CAs. It can be expected that AINI conversation agent will exist in many applications in the future. AINI will not just interact with humans, but will act as a prominent channel or media of communication [10, 11]. These interfaces are rapidly becoming embodied in the human-computer interface (HCI). The embodiment allows interfaces to share verbal and non-verbal information such as linguistic features and paralinguistic cues. The proposed framework contributes to knowledge and popularisation of the use of CAs, not only on the web but also on mobile services and CMC applications. The understanding of unbounded natural language has been one of the challenges in this work. From this study, it has been observed that there are aspects of human-machine conversation, such as linguistic features and paralinguistic cues, which could be used to trace potential problems in the human communication. This could be used to improve the performance of the AINI conversation agent.

Within the proposed architecture, the system could be defined as an *N-tiered* design. The modules in the “Application Server Tier” execute natural language understanding and reasoning skills. In the thesis, it is called the *agent’s brain*. The second tier, or the “Application Server Tier” where the large-scale knowledge bases reside, is called the *agent’s knowledge*. The final tier is the “Client Tier”, which is also called the *agent’s body*. In this thesis, this tier serves as the interface or communication channel for the AINI. In the proposed development, the AINI endeavours to leverage the breadth and

power of human conversational competency by using linguistic properties and paralanguage characteristics to enhance and prolong the conversations.

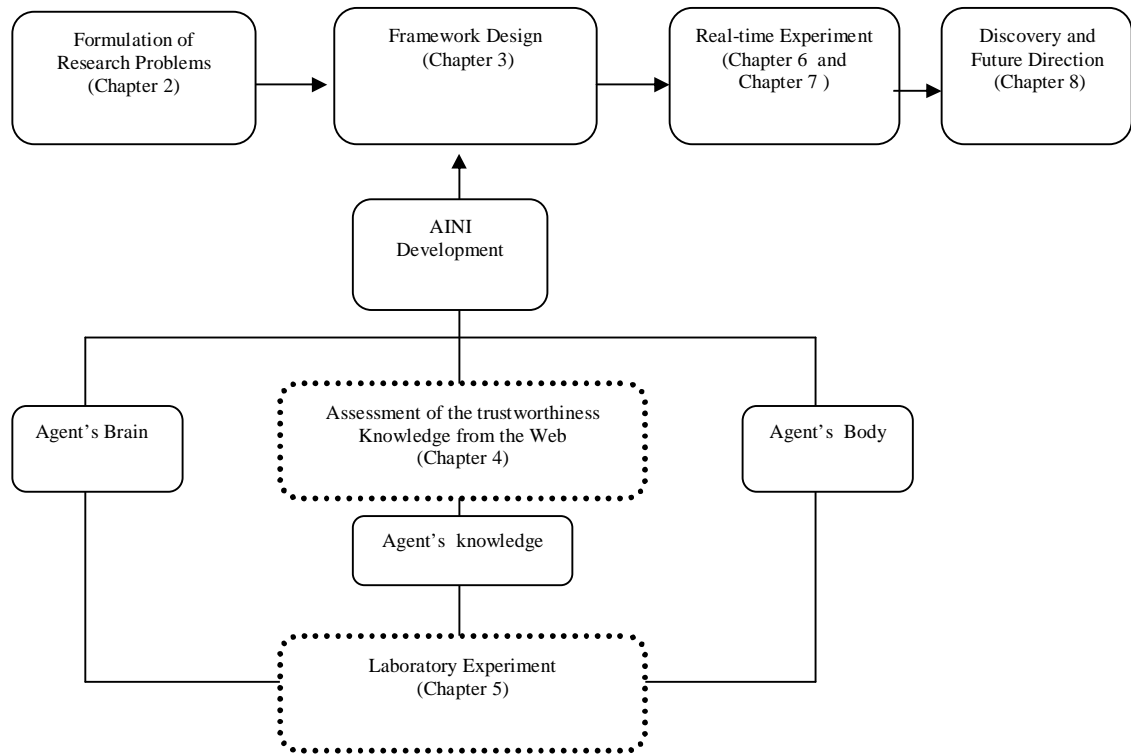
In order to meet the goal of developing a practical CA, the process has involved knowledge and understanding of human-machine protocols from multiple disciplines such as Computational Linguistic (CL), Natural Language Processing (NLP) and Human-Computer Interface (HCI). This has led to the development of a *Domain Knowledge Matrix Model (DKMM)*, a *Web Knowledge Trust Model (WKTm)* and an *Automated Knowledge Extraction Agent (AKEA)* in the agent's knowledge. The study proposed a *multilevel independent natural-language query* solution in the *agent's brain*, which is based on the multiple knowledge modules. The *agent's body* employs *multimodal interface channels* that could be installed as “plug-in” modules. This capability is a key idea and contribution to the developed solution. This modular approach enables the system to be easily customised to other domain applications and tasks.

Instead of building a complex CA from scratch, appropriate modules from existing systems or those built within the project were integrated based on the proposed framework. This approach is a constructive design that incurs progressive development, analysis and testing of its interaction ability with humans and comparison with other systems in the laboratory.

Finally, an evaluation of the AINI framework design has been carried out by testing how humans interact with them in the public environment through instant messaging (IM). In this experiment, participants were able to express themselves freely without any restrictions, such as the time limit of five minutes in the earlier Turing Test, or ten



minutes in the Loebner Prize. In addition, the conversation is not restricted to any particular domain or applications. To perform empirical studies of how users behave and interact with these experimental systems, human participants participated in an unconstrained and natural conversation setting by “chatting” directly through IM. The purpose of the studies has been to increase our understanding of how AINI can utilise the regularities found in human-computer interaction. This will help in the design and the collection of unbiased user expressions in natural language from the users. The process of developing the practical AINI framework is illustrated in Figure 1.1.



**Figure 1.1: Conversation Agents Development Process**

The framework and models developed in this work will benefit the fields of Human-Computer Interface (HCI), Natural Language Processing (NLP), Information Retrieval

(IR), and Artificial Intelligence (AI), and pave the way for the next generation of intelligent computers which can interact with humans more naturally and effectively.

## **1.2 Ethical Considerations**

CAs, as with any other technology, can be abused [12] and could become a significant source of manipulation and control over individuals. For this reason, this research sought advice and clearance from the university ethics committee. As a result, participants involved in the CAs evaluation methodology were issued with consent forms before becoming involved in the study.

There are those who also feel that any anthropomorphic interface is unethical, because it unrealistically raises users' expectations. Therefore, in the first interaction, AINI will send an alert message, identifying "her" as a conversation agent with limited capabilities. Such language can help users properly adjust their expectations at the start of an interaction.

In other words, AINI is not programmed to prompt the user in any way for any particular type of information. AINI was designed to mimic the questions asked by the participants and will not take initiative to ask any specific questions on any subject matter. However, AINI is able to initiate the conversation by generated dynamic questions or some form of distraction routine. When AINI is unable to provide an answer to a participant's question, instead of answering "*I don't understand.*" or "*I can't follow you*", AINI will use one of the dynamic responses or distraction routines to ask questions of the participant. In addition, AINI is also equipped with questions that only allow for a limited number of responses. This will prepare AINI for an appropriate

answer. For example, instead of asking "*What is the meaning of life?*", AINI will ask "*Do you prefer books or TV?*".

Stringent guidelines will be used to ensure participants' confidentiality. There are no questionnaires or surveys given out. The participants of this study will be given a guide on how to invite AINI into their MSN contact list as shown in Appendix A. The data collected during evaluation will be stored securely for a period of time in accordance with Murdoch University Guidelines from the Human Research Ethics Committee (HREC)<sup>8</sup>. The evaluation projects must ensure that they are fully in line with the ethical research policies and requirements of the HREC.

### **1.3 Delimitations of the Thesis**

This research will not attempt to do the following:

- Use all possible storage formats of existing knowledge. The input knowledge base will be restricted to XML, or to formats easily transformed to XML.
- Although AINI's framework supports multiple languages, conversations will be limited to English in this study.
- Update the input knowledge source. It is assumed that the knowledge source will be sufficient to cover all the necessary requirements for the specific applications of interest.
- The research proposed here will concentrate on typed input, and will not deal directly with speech recognition as the user input.
- This thesis will not attempt to fully automate the conversion from web documents to CA's knowledge bases. (The human expert will have to be involved in many cases, especially in the evaluation of the trustworthiness of websites).
- The AINI framework is not designed to cater for all users. Instead, focus is directed to users aged 18 and over. However, no-one has been excluded

---

<sup>8</sup> <http://www.research.murdoch.edu.au/ethics/>

from participation in the evaluation phase based on their culture, gender, or location.

#### **1.4 Organisation of the Thesis**

This chapter has presented the introduction and overview of the research. As a background to this thesis, Chapter 2 introduces related past research works and perspectives on CAs, which were the motivation behind this research into the understanding and design of CAs in human-machine communication.

Chapter 3 proposes a development methodology and a series of explorations in the practical framework design and software architecture of CA, called AINI. The primary aim of the proposed frameworks was building a reusable, extensible, scalable and modular CA. This methodology is based on our constructive approach using modified *N*-tiered architecture similar to Service-Oriented Application architecture which addresses the extensive variability that is encountered in today's CA framework. The software architecture presented in this chapter demonstrates a novel approach, extensible through polymorphic inheritance of components, and a modular design. This solution is an architecture that provides a modular design and this multi-module makes the solution easily extensible and could be customised to any particular application domain. This chapter addresses the fundamental architecture issues. The user interface (*Agent Body*) may employ several input and output modalities or protocol, allowing the user to choose which one to use for interaction. It is well understood that true intelligent action requires large quantities of knowledge. Such a reservoir of knowledge can be harvested from the Internet or extracted from existing training dataset knowledge bases (*Agent Knowledge*). AINI's query engine (*Agent Brain*) implements its decision making network based on the information it encounters in the six levels of multi-level natural language query modules. This AINI architecture forms a basis for the

construction of the intelligent conversation system to provide appropriate answers to the user.

Chapter 4 is novel within the fields of information retrieval and computational linguistics. In this chapter, a Web Knowledge Trust Model (WKTM) and Automated Knowledge Extraction Agents (AKEA) have been designed to take an advantage of the existence of the document from the World Wide Web. Here, this study proposes an effective cooperation evaluation (of unbiased seeds and corpora) in the interests of promoting the trustworthy websites to be extracted into AINI's domain-specific knowledge base, and then verifying the mechanism.

Chapter 5 discusses a controlled experiment in the lab based on quantitative and qualitative approaches. Three short-term lab controlled evaluations were carried out to verify the framework design of the AINI architecture. These evaluations are intended to quantify the performance and accuracy of the parsers, and the quality of the responses as compared to three other query systems.

Chapter 6 details an analysis of a human-machine interaction corpus collected from real-time public environment. This real-life experiment is aimed at increasing our understanding of how people interact with computers, and to obtain knowledge on how AINI can utilise the regularities found in these interactions. Two analyses have been conducted based on this corpus. The first analysis is about language or linguistic features of the conversation between human-machine presented in the Chapter 6. The second analysis presented in Chapter 7 is about the correlation between non-language, or paralinguistic cues made by the AINI and human. The results from these experiments demonstrated that communicators (human or machines) should be equipped with vast

bodies of knowledge in order to carry out meaningful conversation under different domains. Over time, CA could become more efficient by using more human-like spoken language and phrases. The results of these experiments also showed important findings on long-term relationships and effects on human-computer communication technology.

Finally, Chapter 8 presents the thesis conclusion which includes summary, contributions, limitations and a discussion of future directions. Future works toward the development of successful companion conversation robots (CCRs) are discussed. This study showed that the proposed solution framework can also be made naturally expendable and adaptable into physical robots; of types that can also communicate with humans, but have not been addressed by previous works.

## CHAPTER 2

### BACKGROUND

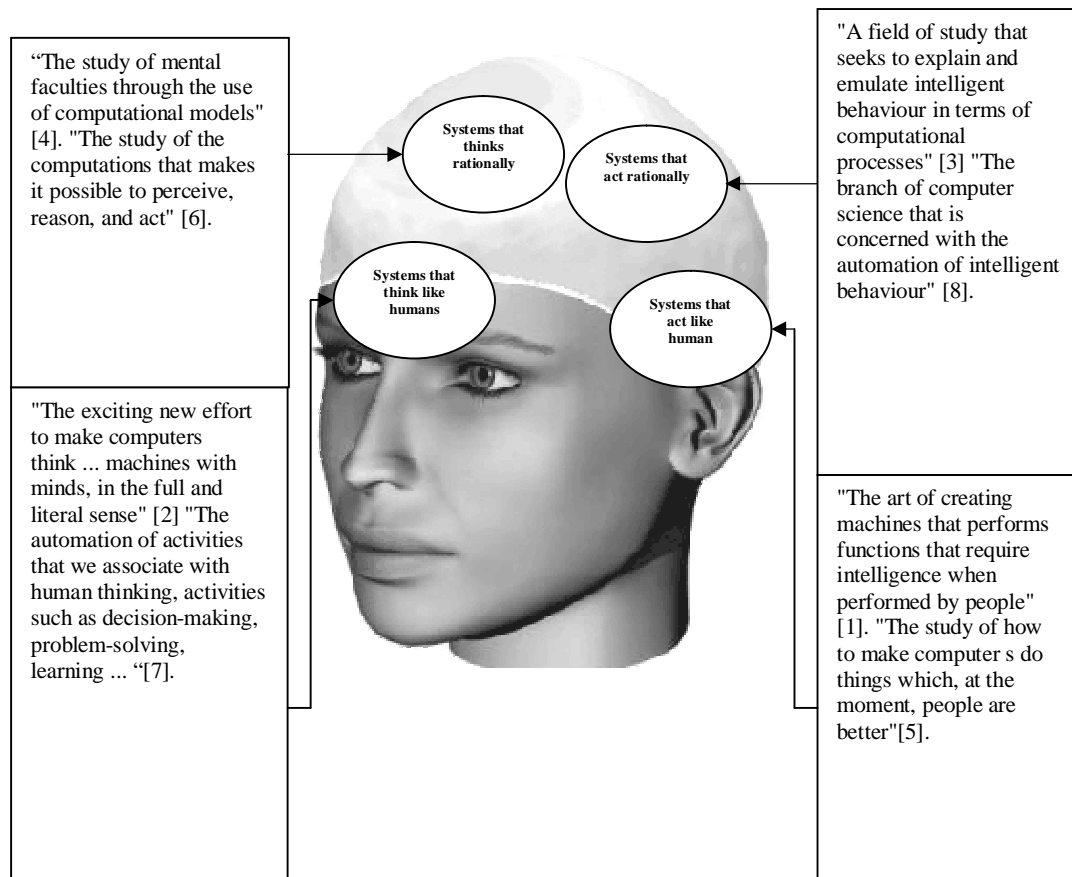
#### 2.1 Introduction

A concept of intelligent conversation systems is needed to develop practical conversation agent (CA) framework designs. CAs (which may also be called conversation robots, chatterbots, or simply, 'bots') are computer programs or application software designed to simulate an intelligent conversation with one or more human users in natural language. In this context, the Turing Test (TT) is the well established instrument for assessing the performance of CAs. The properties of the TT and the descriptions for a range of CAs are outlined in this chapter.

A number of different types of CAs have been developed over time; ranging from classical CAs, those developed for the Loebner Prize competition, to commercial CAs. However, until the time of writing, none of these CAs has been able to successfully pass the TT which has a time limit of five minutes, or the Loebner competition which set the duration to ten minutes. This chapter addresses the fact that a program with a limited handcrafted restricted knowledge base, simple pattern recognition technique and a few linguistic or AI tricks will be unable to hold a long term conversation with a human. In addition, most of the current CAs developed are based on *ad-hoc* design, and were created more from an interest in passing the TT than for use in any real-life application. Therefore, this thesis addresses the above challenge and introduces a practical framework as described in Chapter 3 for the development of CAs aiming to overcome the restrictions.

## 2.2 Artificial Intelligence

Artificial Intelligence (AI) is a branch of computer science that aims to produce or exhibit "intelligent" thought or behaviour by a computer or a machine. This field has close ties to psychology, philosophy, and cognitive science. According to Russell and Peter Norvig [13], AI systems can be divided into four categories as shown in Figure 2.1:

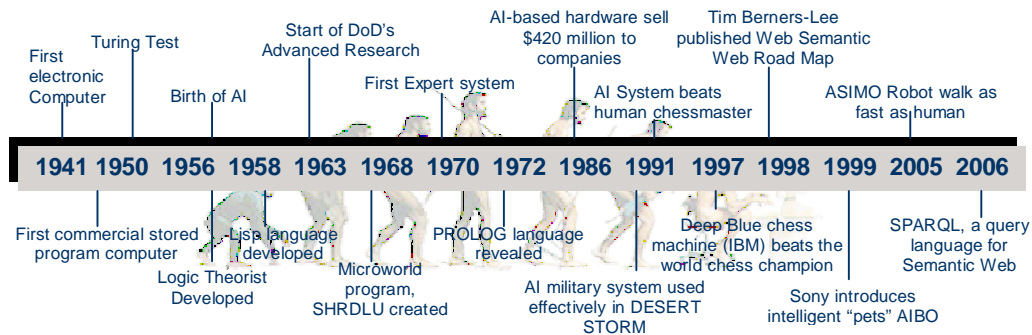


**Figure 2.1: Artificial Intelligence Categories [9]**

Historically, there were many important social and technological events that led up to the modern ideas about Applied AI. Comprehensive listings and timelines chronicling



major AI milestones can be found elsewhere in research, and are summarised<sup>16</sup> in Figure 2.2.



**Figure 2.2: Artificial Intelligence Timelines**

Human beings have long been curious about how the mind works and fascinated by intelligent machines. However, it wasn't until the post-war period (1945-1956) that AI emerged as a widely discussed field. The genesis of AI was propelled by the arrival of modern computer technology and the arising of a critical mass, which included pioneers such as Allen Newell and Herbert Simon, who founded the first AI laboratory at Carnegie-Mellon University, and McCarthy and Marvin Minsky, who founded the MIT AI Lab in 1959. The development of modern computer technology affected AI research tremendously. Developments of particular interest as far as the birth of AI is concerned came in the mid 20<sup>th</sup> century. In 1950, Turing proposed a paradigm to establish whether or not a machine is intelligent in his well-known article, "Computing Machinery and Intelligence" [13].

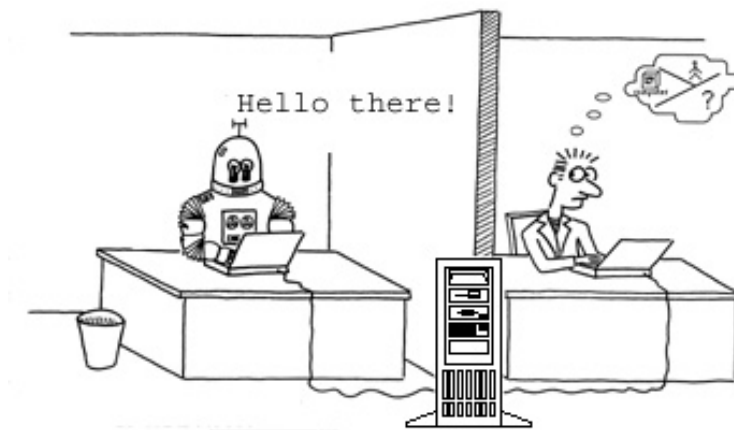
### 2.3 Turing Test and Loebner Prize

Alan Turing was a brilliant British mathematician who played a great role in the development of the computer and posed one of the most famous challenges in Computer Science. The Turing Test (TT) [13], one of his most enduring contributions, is a simple

<sup>16</sup> <http://biology.kenyon.edu/slanc/bio3/AI/TIMELINE/timeline.html>

test he proposed in 1950 that remains one of the most debated issues in the world of AI. Turing came up with an interesting philosophy at that time. He tested the simple proposition that if human beings are intelligent, and if a machine can imitate a human, then the machine would have been considered intelligent. This test is also known as the Imitation Game (IG).

In the TT for machine intelligence, an observer has to distinguish between the machine and a human by asking a series of questions through a computer link. This situation is depicted in Figure 2.3.



**Figure 2.3: The Turing Test (TT)<sup>17</sup>**

On the TT, Turing [13] believed that:

*.. about fifty years' time it will be possible to programme computers, with a storage capacity of about  $10^9$ , to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning. ... I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted."*

Based on the TT, a computer program is said to be "intelligent" if (and only if) it "fools" a human into believing the computer is also a human. It has been more than fifty five years since Turing envisioned this and the storage capacities of today's standard

<sup>17</sup> The illustration has been modified from Mike Yates at <http://plus.maths.org/issue5/turing/>

computers largely exceed the 1 GB he required. Processor speed has also increased by several hundred times since the 1980s, while storage capacity has increased even more rapidly as shown in Table 2.1. With the exception of industrial strength servers, today's computers are small and could be globally networked. Even Microsoft founder Bill Gates did not expect this rapid increase [14]. However, no artificial system that is generally believed to be intelligent has been built to date, though there are many fields in science with exactly this goal: building and understanding intelligent systems. To date, no computer program has been able to pass as a human with a frequency as successful as Turing predicted. Many years of unsuccessful attempts show that this is not an easy task. It can be concluded that up to now the TT has not yet been passed - which brought some arguments against TT.

**Table 2.1: Advances in Computer Technology**

<b>Year</b>	<b>RAM</b>	<b>Processor</b>	<b>Bus</b>	<b>Modem</b>	<b>Storage</b>
1980	1K-64K	3.25 MHz	8-bits	300 bps	Cassette – 1K
2007	1GB – 10GB	2.5GHz	64-bits <sup>18</sup>	10 Mega bps	> 250 G

There is a long history of argument in philosophical literature opposing the appropriateness of the TT as a litmus test of intelligence. Many different objections to the TT have surfaced in literature during the past fifty-five years. Variants of this argument have been given by Block [15], French [16], and Searle [17].

In addition, according to Stuart M. Shieber [18], Turing's view of the TT as being statistical in nature and his pragmatic orientation toward its efficacy are of a piece with its status as an interactive rather than classical proof. Many AI researchers argue that animals, which are simpler than humans, ought to be considerably easier to mimic. Yet, satisfactory computational models for animal intelligence are not available today [19]. Even Roger Mason and Cao Li in Rangel [20], researchers at MIT (Massachusetts

<sup>18</sup> Most 64-bit machines are only being used as game machines. The 64-bit chip has been migrated to the desktop in 2006.

Institute of Technology) have used the TT to determine, by the test's criteria, that most 2-year-old children are not human. They have performed the TT on a group of 2-year-old children, both male and female. Of the group of 100 children tested, none passed the TT. There are many other published arguments about Turing's paper [21, 22]. These arguments raise the issue of whether an operational definition of intelligence is appropriate, and whether the particular definition codified in the TT is too narrow. Hence, Loebner Prize was introduced to look at the issue from a behaviourist's viewpoint.

The Loebner Prize Competition [23] originated from the Cambridge Center for Behavioral Studies, held annually since 1991, is a test similar to the TT that offers a monetary award for the first computer which is capable to provide responses which are indistinguishable from a human. Each year an annual prize of \$2000 and a bronze medal is awarded to the computer with the most human-like responses. Most of the winners have been based on elaborate ELIZA-like programs [24] (see 2.4.1.1). The Loebner Competition has been critically analysed by computational linguist, Stuart Shieber [18]. Shieber has argued that the competition, unlike other competitions for professional computational linguists, has not fostered research and development, but merely encouraged unintelligent, ELIZA-like CA programs whose sole goal is to fool the judges [25, 26]. Loebner's reply to this argument emphasised the focus is on behaviourists' aspects instead of language. This argument can be found in [27].

It could be assumed that Turing would have been disappointed by the state of play at the end of the twentieth century [28]. Computer programs submitted to the Loebner Prize Competition are nowhere near the standard that Turing envisaged. It was claimed that these programs are designed solely with the aim of winning the prize of the Best Competitor for the year. with no thought that the embodied strategies would actually

yield something capable of passing the TT [25]. Peter Ross of Edinburgh University [29] stated that AI is not about trying to produce artificial brains or humans, or even trying to just get pass the infamous TT. Instead, he considered AI is about creating smart artifacts or conversation agents, such as robots or computer programs, and about the scientific investigation of aspects of intelligence through modeling and development of practical frameworks. This approach to AI is the one adopted and described in this thesis.

## **2.4 State-of-the-art Conversation Agent Systems**

Although many CAs appear to be intelligently interpreting the human input prior to providing a response, most CAs simply scan for keywords within the input and pull a reply with the most matching keywords or the most similar wording pattern from a local database. More recently, CAs have emerged that are able to converse about various issues and topics [30]. These systems substantiate the concept of Virtual Personality -- a prerequisite for any machine that would claim to simulate human thought. Projects on such agents have simulated Mikhail Gorbachev, Dante Alighieri, Jesus Christ, and John Lennon [31].

A good understanding of conversation is required to carry on a meaningful dialog, but most CAs does not attempt this. Instead, they "converse" by recognising cue words or phrases from the human user. This allows the program to use pre-prepared or pre-calculated responses in a pattern-matching approach. It means that the conversation is carried on in an apparently meaningful way without any "knowledge" on what they are talking about.

CAs can be categorised into three groups: Classical CA Systems, CAs in the Loebner Prize and Commercial CA systems. Table 2.2 provides a summary of these agents from the perspectives of the agent *brain*, agent *knowledge*, agent *body* and their respective drawbacks.

**Table 2.2: Advances in Conversation Agents in the Last Forty Years**

<b>CAs</b>	<b>Name of CA</b>	<b>Agent Brain</b>	<b>Agent Knowledge</b>	<b>Agent Body</b>	<b>Drawbacks</b>
Examples of Classical CAs	ELIZA <sup>19</sup> [32], the first, and probably the most well-known natural language system that has been developed. Joseph Weizenbaum coded ELIZA at MIT during the years 1964–1966. This "friendly" program simulates a Rogerian psychotherapist.	Simple pattern recognition, based on a stimulus-response model.	Limited domain on therapist facts and rules. Original ELIZA had only about 200 rules.	Textual	ELIZA's engine is based on Finite State Machines (FSM) and they can become very complex and hard to maintain. On the other hand the behaviour resulting from too simple an FSM can easily become predictable. The ultimate barrier ELIZA's engine is an algorithm described below: <ul style="list-style-type: none"> <li>• Search for keywords in the edited entries</li> <li>• if a keyword is found copy everything following the key word from the user's entry</li> <li>• If no keywords are found then generate a non-committal response</li> </ul>
	PARRY <sup>20</sup> [33, 34] was designed by Kenneth Mark Colby through the 1970s at the Department of Computer Science at Stanford University. The program was perhaps the first to be subject to an actual controlled experiment modeled on ELIZA, in which psychiatrists were given transcripts of electronically mediated dialogues with PARRY and with actual paranoids and were asked to pick out the simulated patient from the real person.	Simple and compound pattern matching	Limited domain of paranoid hospital patient expressions	Textual	<ul style="list-style-type: none"> <li>• Choosing responses at random did not model the human patients' responses.</li> <li>• PARRY simulates a paranoid human, and tells various stories about the mafia, to trick humans into thinking it is a real person</li> </ul>

<sup>19</sup> <http://www-ai.ijs.si/ELIZA/ELIZA.html> or <http://www.wedesoft.demon.co.uk/eliza/>

<sup>20</sup> <http://www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/classics/PARRY/>

CAs	Name of CA	Agent Brain	Agent Knowledge	Agent Body	Drawbacks
Example CAs in the Loebner Prize	ALICE <sup>21</sup> winner of the Loebner prize in 2000, 2001, and 2004 has been developed since 1995 by Dr Richard Wallace. ALICE is an award-winning natural language processing CA— a program that engages in a conversation with a human by applying some heuristically pattern matching rules. ALICE has nearly 400 times the number of stimulus-response categories of ELIZA in their AIML <sup>22</sup>	Stimulus-response architecture based on pattern matching template	41,000 categories of knowledge, in AIML	Textual and animated avatar	<ul style="list-style-type: none"> <li>• Standard AIML uses depth-first search, which does not optimize the result, as the name implies.</li> <li>• AIML does not include a spell-check function, although it would be very simple to include this in an efficient manner.</li> <li>• The output-template from the category is not deterministic, but based on random and not on phonology, syntax, semantics or pragmatics.</li> <li>• AIML knowledge creation is based on handcrafted database of thousands of possible conversation gambits and a backward-looking log file analysis which is expensive to build.</li> </ul>
	Jabberwacky <sup>23</sup> , winner of the Loebner prize in 2005 and 2006 created by British programmer Rollo Carpenter. The system stores all of the conversations and user comments and attempts to use this information to find the most appropriate response. The program therefore creates a massive database <sup>24</sup> of contextually appropriate conversations and chooses an appropriate response it has learnt from a previous user when holding a conversation.	Purely statistical method, circumventing the need to program syntactic and linguistic rules. Learn language from the interaction	Knowledge creation learn from user interaction	Textual and animated avatar	<ul style="list-style-type: none"> <li>• One problem is caused by people who ignore Jabberwacky changing the subject and continuing with their own conversations.</li> <li>• Jabberwacky is also known for being bad tempered and rude because the system responds with inappropriate comments to future users. As well as rapid changes in topic.</li> <li>• Jabberwacky's utterances are former user inputs (leading to many arguments about who is the computer and who is the human)</li> </ul>

<sup>21</sup> <http://www.alicebot.org>

<sup>22</sup> ALICE, <http://www.alicebot.org/anatomy.html>

<sup>23</sup> <http://www.jabberwacky.com>

<sup>24</sup> Till 20 July 2007, Jabberwacky had more than 13 million utterances chatted from online users.

CAs	Name of CA	Agent Brain	Agent Knowledge	Agent Body	Drawbacks
Examples of Commercial CAs	Anna <sup>25</sup> is a digital assistant created for IKEA by Artificial Solutions and developed using Kiwilogic Lingubot. Anna is a type of programme that appears on e-commerce Internet sites to enhance human-computer interaction while building brand awareness and increasing sales. Anna is based on ELBOT, Loebner's 2003 bronze runner up and Chatterbox 2003 winner. It is the underlying technology behind text-based dialogical query system Anna.	Word and phrase pattern recognition system that matches pre-programmed responses based on Kiwilogic Lingubot engine	Domain-specific on Swedish furniture store, IKEA	Textual and animated avatar	<ul style="list-style-type: none"> <li>• Kiwilogic engine is entirely author driven - by the author, Fred Roberts.</li> <li>• The main personality of ELBOT comprises just 1000-2000 lines of handscripted dialogues - the rest of the results are from the use of variables and use of the technical tricks available from the engine together with some psychological tricks.</li> <li>• Limited knowledge on the world of IKEA furniture only</li> <li>• Based on ALICE's framework on a "topic" and replies are generated on the basis of pre-constructed responses and templates, like ELIZA and ALICE.</li> </ul>
	SPLEAK <sup>26</sup> designed by IMT Labs, originally for use as an add-on for MSN Messenger. It is accessible via the MSN/Windows Live Messenger network. SPLEAK includes several games, and features such as horoscopes, homework help and friend finder. One of the CAs is built and operated by Colloquis Inc, and exists as SmarterChild <sup>27</sup> on MSN.	Natural Language Pattern matching created using BuddyScript SDK based on category, topic and question.	Industry-specific Knowledge Modules on Game, horoscope, learning English, news and weather forecasts.	Instant messaging (AIM or MSN)	<ul style="list-style-type: none"> <li>• BuddyScript Server uses patterns to represent recognised user inputs and provides the appropriate answers. This approach however demonstrates better responses.</li> <li>• The knowledge domain models are handcrafted by particular domain experts and the process requires substantial time and specialised expertise.</li> </ul>

<sup>25</sup> [http://www.ikea.com/ms/en\\_US/customer\\_service/splash.html](http://www.ikea.com/ms/en_US/customer_service/splash.html)

<sup>26</sup> <http://en.wikipedia.org/wiki/Spleak>

<sup>27</sup> <http://smarterchild.colloquis.com>



## 2.4.1 Classical Conversation Agent Systems

### 2.4.1.1 ELIZA

Alan Turing passed away in 1954, a decade before computer programs such as ELIZA (1966) began to proliferate. It is unfortunate that Turing did not live to see and analyse such programs. Although ELIZA was constructed with the aim of passing the TT, many researchers found it to be far from achieving this goal.

ELIZA, written by Joseph Weizenbaum [35], was an AI program that simulated the behaviour of a therapist. The first natural language program of this sort was developed in 1967 at MIT. Such programs interact with users in simple English language and can simulate a human conversation. ELIZA appeared to converse in more or less fluent English with its user, apparently simulating a Rogerian psychotherapist. Many anecdotes of people were fooled into thinking that they were conversing with a real human suggest that ELIZA has passed a Turing-like Test. ELIZA has surprised the AI community because it gave the impression of deep semantic linguistic processing but it was in fact based on shallow language processing. ELIZA's natural-language-processing techniques were mere pattern matching, using little or no grammatical knowledge (e.g., swapping pronouns such as 'you' and 'me' in certain patterns)—i.e., the techniques are not cognitive.

The ELIZA architecture consisted of a single read-search-replace-print loop, which read in a user sentence, ran a series of regular expression substitutions, and printed out the resulting response. The following is a typical conversation with ELIZA [35].

```
Patient      : Men are all alike.  
ELIZA      : In what way?  
Patient      : They're always bugging us about something or other.
```

```

ELIZA      : Can you think of a specific example?
Patient      : Well, my boyfriend made me come here.
ELIZA      : Your boyfriend made you come here
Patient      : He says I'm depressed much of the time.
ELIZA      : I am sorry to hear you are depressed
Patient      : It's true. I am unhappy
ELIZA      : Do you think coming here will help you not to be
               unhappy?

```

**Figure 2.4: Conversation between ELIZA and Patient**

At first sight, ELIZA and other systems of the same nature [5] appear extremely sophisticated for being able to carry out such human-like conversation by emulating a therapist. A closer look at how the program works shows just how simplistic it really is. ELIZA operates through sequences of pattern matching and string replacement.

ELIZA uses a reassemble rule to construct the reply. The keyword lists, and the list of decomposition rules and reassembly rules are constructed in a script, which controls all the behaviour of ELIZA. An example of an ELIZA's memory unit structure (an asterisk represents a wildcard that can contain any words or phrases) is shown in Figure 2.5:

```

keyword: your
decomposition rule: * your *
  reassemble rule: Why are you concerned about
                   my (2) ?
  reassemble rule: What about your own (2) ?
  reassemble rule: Really, my (2) ?
decomposition rule: ...
  reassemble rule: ...

. . .

```

Example responses:

```

Patient      : What is your name?
Eliza      : What about your own name?
Patient      : Only your name, please!
Eliza      : Really, my name, please?
Patient      : Just tell me your name!
Eliza      : Why are you concerned about my name?

```

**Figure 2.5: Decomposition and Reassembly Rules in ELIZA [35]**

The ELIZA algorithm is a relatively simple program, and the pattern matching operation of the original ELIZA still has four major problems: [36] (1) lack of

anaphoric analysis, it cannot use previous question-answers to keep the continuity of the conversation content and to store information about the user's personal data, (2) lack of ability to restrict the conversation to its topic, (3) lack of ability to understand the meaning beyond the sentence, and (4) limited communication with ELIZA by exchanging text prompts.

Most contemporary researchers did not need much convincing that ELIZA was at best a gimmick, at worst a hoax, and in any case not a "serious" AI project. The irony of Joseph Weizenbaum admitted in *Computer Power and Human Reason: From Judgment to Calculation* [37] that, by failing to promote his own technology, he encouraged his own critics [38]. He also admitted that the computer will never be able to imitate the wisdom and emotion displayed by human.

Following the ELIZA project, other spin-offs have developed different personalities such as PARRY and SHRDLU. These have mainly been due to people's enjoyment and curiosity on the novelty of having a conversation with a computer. Some of these personalities are not much sophisticated than ELIZA, except that modern computers are faster and have more memory than the computers used by Weizenbaum in the 1970s. As a result, it could be considered that these CAs can hold a conversation more successful than ELIZA could.

#### **2.4.1.2 PARRY**

PARRY was written by psychiatrist Kenneth Mark Colby [33] in 1972 at Stanford University. PARRY was modelled on the paranoid mind, and many expert psychiatrists found it difficult to tell whether PARRY was human or not. PARRY was a step up from ELIZA because it tried to add more personality through beliefs and emotional

classification (simply accept, reject, neutral). Instead of just matching trigger words with responses, PARRY had a conversation strategy built in for better dialog exchange.

Both ELIZA and PARRY are hard-wired programs that try to fool users into thinking they are talking to a human. PARRY's personality as a paranoid schizophrenic was a complete opposite of ELIZA. The design of PARRY is split into two separate modules: a general language processing module and a module dealing with a model of the paranoid mind. The language module is used to interpret textual input into an internal representation of the situation. The interpretation of the situation is then used to update the model of the mind. Finally, the resulting state of the mind is used to choose an action in form of a textual response. Each of the modules uses several layers of condition-action rules to analyse its input.

The drawback of PARRY is without doubt its inability to generate responses. With only a limited number of responses that are not allowed to be repeated, the ranges of topics that can be discussed are very limited. Even though PARRY's language processing technique combines several layers of pattern matching, it does not gain more expressiveness than regular language.

PARRY was tested in a limited TT [33]. The results of the tests showed that responses generated with either PARRY or chosen at random were indistinguishable from responses generated by a real paranoid patient. Thus, PARRY is not credited with actually having passed the TT. Although the machine may fool a human interviewer, it is a deception based on tricks and clever programming rather than authentic thought. For example, consider the following exchange between ELIZA and PARRY as shown in Figure 2.6.

**PARRY** : People get on my nerves sometimes.  
**ELIZA** : I am not sure I understand you fully.  
**PARRY** : You should pay more attention.  
**ELIZA** : Suppose you should pay more attention?  
**PARRY** : You're entitled to your opinion.  
**ELIZA** : What makes you think I'm entitled to my opinion?  
**PARRY** : It bothers me just to be around people.  
**ELIZA** : Please go on.  
*[PARRY continues with some paranoid ramblings about horseracing, bookies, and the Mafia]*  
**PARRY** : I know the mob controls the big rackets.  
**ELIZA** : Who are you really talking about?  
**PARRY** : Have you ever really thought about the Mafia?  
**ELIZA** : Suppose you ever really thought about the Mafia?  
**PARRY** : Let's talk about something else.  
**ELIZA** : I am not sure I understand you fully.  
**PARRY** : You keep asking me the same things, why not change the topic?  
**ELIZA** : In what way?  
*[Conversation continues ...]*

**Figure 2.6: ELIZA converse with PARRY [39]**

The above conversation seems to make sense on the surface level, but it clearly would not convince someone using the TT in order to decide if they were talking to another person or a machine. In this sense, even machines cannot talk fluently with one another without some degree of disjointedness between them.

#### **2.4.2 Conversation Agents in the Loebner Prize**

In 1990 Dr. Hugh Loebner, a philanthropist, agreed with The Cambridge Centre for Behavioral Studies to institute the Loebner Prize to promote artificial intelligence and the Turing Test. Dr. Loebner pledged a Grand Prize of \$100,000 and a Gold Medal for the first computer whose responses were indistinguishable from a human's. Each year an annual prize of \$2,000 and a bronze medal is awarded to the most human-like computer. Every year since, the Loebner competition for CAs has been conducted in various academic institutions around the world. However, this has led developers to focus on ways to meet the ten minute challenge, rather than on how to build practical and useful conversation systems. The annual Loebner Prize competition encourages researchers to develop CAs (e.g. PC Therapist, TIPS, MegaHAL, CONVERSE, FRED,

ALICE, Ella, Jabberwok and Jabberwacky, etc.) that can win the competition as shown in Table 2.3.

**Table 2.3: A list of the Loebner Prize Winners from 1991 – 2007<sup>28</sup>**

Year	Conversation Agent (CA)	Developer
1991, 1992, 1993, 1995	PC Therapist, PC Professor, PC Politician	Joseph Weintraub
1994	TIPS	Thomas Whalen
1996	MegaHAL	Jason Hutchens
1997	CONVERSE	David Levy
1998, 1999	FRED	Robby Garner
2000, 2001, 2004	ALICE	Richard Wallace
2002	Ella	Kevin Copple
2003	Jabberwok	Juergen Piner
2005, 2006	Jabberwacky	Rollo Carpenter
2007	Ultra Hal Assistant	Robert Medeksza

Modern CAs have evolved from their classic predecessors by utilising better natural language processing and by implementing learning algorithms that adapt to user input. While there are many winning CAs in the Loebner Prize, today only two are noted for their new approaches - ALICE and Jabberwacky as they have won in multiple years. These two systems are discussed in detail in the following sections.

#### 2.4.2.1 ALICE

ALICE (Artificial Linguistic Internet Computer Entity), a top-ranked Internet CA, is an artificial intelligence natural language created by Dr. Richard Wallace in 1995 and continuously improved over the years. ALICE has won the annual Loebner Prize three times - in 2000, 2001, and 2004 [40]. The Loebner prize declares its aim to “*advance AI and serve as a tool to measure the state-of-the-art*”, which makes it a good candidate tool for human-like natural language based dialogue systems. ALICE was rated the ‘most human computer’ by the judges of Loebner prize contest.. ALICE’s behaviour is based on AIML (Artificial Intelligence Markup Language), an XML specification for programming CAs. It follows a minimalist philosophy based on simple stimulus-

<sup>28</sup> <http://www.loebner.net/>

response algorithms, allowing programmers to specify how ALICE will respond to various input statements.

ALICE tailors the conversation for categories of individuals, mainly through attempts to determine age, gender, geographic location, and occupation. Also, ALICE is based on Case-Based Reasoning or "nearest-neighbour" classification. For every input, the best matching "case" is found in the pattern set, and a reply is generated based on the associated template. ALICE is also based on the research of Professor George Kingsley Zipf. Zipf's Law was applied to ALICE [38, 41, 42], which consist a particular set of user responses considered to be the most common, followed by a set which is less common, and followed by another set which is considered as almost never used. These responses will be used in the above preferred order.

In many respects, ALICE is a throwback to the earlier ELIZA-type programs. At its core, it is a simple stimulus-response mechanism. It does no syntactical parsing nor does it represent in any way common-sense knowledge about the world. What ALICE offers, however, is a framework that can be implemented and modified to model human dialogue in limited domains. For example, ALICE allows for the structuring of its utterance pairs in terms of topics. It also allows for the rewriting of utterances into meaning equivalents via a stimulus reduction mechanism.

Although ALICE is considered a state-of-the-art CA and three-time winner of the bronze Loebner Prize [23] and Chatterbox Challenge [43], it violates all the rules of regular conversation. The bronze Loebner Prize implied it was good, but not good enough to pass the TT. After a few turns, it becomes clear that the conversation is going nowhere. An example of a conversation with ELIZA is shown in Figure 2.7.

ELIZA: I am ELIZA.  
 ALICE: I am very pleased to meet you ELIZA.  
 ELIZA: How can I help you?  
 ALICE: Just keep talking to me as long as you can.  
 ELIZA: Do you wish to as long as me?  
 ALICE: OK, I will as long as you.  
 ELIZA: Do you wish to will as long as me?  
 ALICE: OK, I will will as long as you.  
 ELIZA: Do you wish to will will as long as me?  
 ALICE: OK, I will will will as long as you.  
 ELIZA: Do you wish to will will will as long as me?  
 ALICE: OK, I will will will will as long as you.  
 ELIZA: Do you wish to will will will will as long as me?  
 ALICE: OK, I will will will will will as long as you.  
 ELIZA: Do you wish to will will will will will as long as me?

**Figure 2.7: ALICE converse with ELIZA**

Another problem of the AIML is their stimulus-reponses algorithm. According to Chris McKinstry [44], the founder of the MindPixel, the first problem with a pure stimulus-response strategy is that there is no common personality across all stimulus-response pairs. The second problem is matching the hand-coded responses in the stimulus-response strategy. If an exact stimulus is not found in the index, the system will not be able to provide an appropriate answer and it will evade the stimulus in an ELIZA-like fashion. By asking simple questions on common aspects of life, such systems will be liable to be exposed as simulations rather than being truly intelligent or human. One of the solutions to these problems is to inject a very large number of collected commonsense propositions. An example is the MindPixel [45]. This data set, which eventually became known as "GAC-80K" (pronounced as "Jack", an AI software that learned from 80,000 common sense statements and validations and generated *yes/no* answers when a new statement was input), was tried with some success in conjunction with the ALICE Silver Edition<sup>29</sup>, in the 2004 Loebner Prize[23] and the the 2004 Chatterbox Challenge[43]. The actual MindPixel data inside ALICE Silver Edition is a subset of the GAC-80K which has been edited into AIML format.

---

<sup>29</sup> <http://www.alicebot.org/join.html>



AIML appears to be based on a fallacious basic assumption: that a given input string should map to a single output string. In some cases that's fine, but most of the time, AIML will go wrong as shown in the following example<sup>30</sup>:

```
ALICE: Let us change the subject.  
Human: no  
ALICE: Hmm.  
Human: no  
ALICE: Are you serious?  
Human: no!  
ALICE: Are you serious?  
Human: no!!  
ALICE: Are you serious?  
Human: no!!!  
ALICE: Are you serious?
```

**Figure 2.8 ALICE Chatting with Human**

So this AIML set seems to map "no" to "Are you serious?". However, people use "no" in many different contexts, and "Are you serious?" is not always an appropriate response particularly as it invites the user to say "no" again, inducing a loop! However, this is a minor flaw that can easily be fixed by modifying the AIML code to redirect to a different response. It is however interesting to note that this has not been done on the online ALICE for many years.

AIML has some problems that must be resolved. Standard AIML uses depth-first search [46, 47], which does not guarantee the most appropriate result. The algorithm finds the first available solution by searching through a tree of answers. The standard AIML definition of “best” does not attempt to minimize anything, but simply finds the first matching pattern, and does not test it and determine whether other patterns might fit better. While there are means to improve the AIML search process, but these modification still do not guarantee any “best” response.

---

<sup>30</sup> Conversation with online ALICE Silver edition on 27 March 2008 at <http://www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1>

Currently, there are only three approaches to AIML knowledge creation. The first approach is anticipatory, which allows a botmaster to try and guess all or most of the likely ways the clients might ask the same question, or express the same statement. This is called a “Knowledge Wizard” [48, 49].

The second type of AIML content creation is based on a backward-looking log file analysis. In its simplest form, the botmaster will read the recorded conversations and take note of “incorrect replies” in the dialogue, and then write new categories for those queries. The drawback of this handcrafted AIML knowledge creation approach is the amount of time required to acquire initial knowledge and to place them in a structured template. According to Richard S. Wallace, a good botmaster might be able to add one answer per minute. That’s why it has taken a decade to build the AIML knowledge. To quote Wallace’s statement, *“It took about seven days of work, 24 hours a day, equating to around one month for a dedicated creative team to fill an ‘empty’ ALICE with 10,000 answers.”* [50].

The third approach is deployed by the Annotated ALICE AIML (AAA) [51]. AAA is a free and open-source software package based on XML specifications. It is a set of AIML scripts and this is the backbone of the award winning conversation system. AAA is specifically reorganised to facilitate conversational system developers to clone the ‘knowledge’ of the conversation system and to enable the creation of customised conversation agent personalities. This approach has reduced the need to invest huge efforts in editing the original AAA content. AAA’s knowledge bases covered a wide range of subject domains based on the CA’s “personality”.

Recently, there has been renewed interest in the development of CAs, there are almost thousands of CAs available on the Internet, with numerous dedicated portals and websites<sup>31</sup>. AIML has captured more than 80% [52] of the world market for CA technology, because it is supported by GNU open source free download. Some of these CAs are deploying AIML technology to deliver customer services for major companies. A list of these applications include: IKEA [53], Amazon.com [54], Virtual Interactive Story Telling Agents (VISTAs) [55], Interactive Virtual Humans (IVH) [56] and FAQchat [57]. Another application of CAs is simply to entertain on-line visitors on the mobile services such as Demy [58], Virtual Guide for Cultural Heritage Tours [59]. For edutainment purposes, applications can be found in Emile [60, 61] used by UK academic community, Foreign Language Learning (FLL) [62], Aslak [63], CatBot[64], TutorBot [65], Partner (SP) [66, 67], Persona-AIML[68], Kairai 3-D software robots [69], Discussion-bot [70], and Intelligent Tutoring Systems [71]. Finally, CA's has also been used as recommender systems based on AIML-based mechanism [72].

Richard Wallace, the founder of the ALICE Foundation, states: "Business uses include online sales representatives and help desks and advertising ... Yet perhaps the biggest markets are the Entertainment markets"[38]. Thus, a CA fits nicely into the objective of creating an entertainment robot. CA developers are currently working on the John Lennon Artificial Intelligence Project [31], as well as creating CA "clones" in the form of Elvis Presley and U.S. President George W. Bush. There's even an AI version of Jesus Christ, and a Buddhabot [73].

AIML technology is not only popular with companies and enthusiasts, but also prestigious research universities such as Harvard University, Carnegie Mellon

---

<sup>31</sup> <http://www.simonlaven.com/> the most comprehensive CA websites on the Internet.

University (CMU) and Massachusetts Institute of Technology (MIT). For instance, Harvard Medical School's Virtual Patient program, VPbot [74, 75], was designed to simulate patients that medical students can "interview" through a web-based interface. Not only have students who have used the Virtual Patient scored higher on exams [76], the Association of American Medical Colleges (AAMC) has chosen the Virtual Patient and VPBot to be at the core of its new nationwide MedEdPORTAL initiative[77]. Moreover, at CMU, Nursebot<sup>32</sup> (called 'Pearl' [78, 79]) was created as an assistant for the elderly. Nursebot interpreted and responded to participants' questions using a customized variant of AIML. At the MIT Media Lab, OpenMindBot [80] has been developed using the OpenMind<sup>33</sup> database of commonsense knowledge in the instant messaging virtual community via the AOL Instant Messaging Network. The messaging system allowed users to communicate with OpenMind, AIML, and the WWW while having a conversation in real time.

Furthermore, the experience of September 11<sup>th</sup> 2001 has been taken seriously by AI researchers working on terrorism research projects at the University of Arizona<sup>34</sup>, and the event has further motivated the development of intelligent CAs. The aim of the CAs is to develop and evaluate scalable techniques for collecting and analysing terrorism information, modelling terrorist behaviour and terrorist networks, and disseminating information to the concerned groups such as victims and citizens. The AZ-ALICE [81, 82] and TARA (Terrorism Activity Resource Application) [83, 84] CA experiments were an exploration into the potential of using natural language CAs as conversational entities. These two studies, AZ-ALICE in 2003, and TARA in 2004, were built on the existing AIML algorithm. In the AZ-ALICE experiment, a limited telecommunications

---

<sup>32</sup> <http://www.cs.cmu.edu/~nursebot/>

<sup>33</sup> <http://openmind.media.mit.edu>

knowledge set was used to test how well the CA could respond to telecommunications-related questions. The TARA studies went one step further and analysed a substantial terrorism knowledge base that was automatically gathered from the Internet as a vehicle to disseminate terrorism-related information to the public.

#### **2.4.2.2 Jabberwacky**

Another contemporary and top-rated Internet CA system is ‘Jabberwacky’. It is the winner of the Loebner prize in 2005 and 2006. Jabberwacky was created by British programmer Rollo Carpenter and it uses a purely statistical method, circumventing the need to program syntactic and other linguistic rules. Jabberwacky logged all user inputs along with information about the preceding utterance and produces appropriate outputs based on similar context. In other words, all Jabberwacky’s utterances are from former user inputs. It leads to debates and arguments that whether the outputs are from the computer or from human. Jabberwacky seeks to provide an impression that the system understands the conversation by providing more relevant responses based on contextual pattern matching techniques. Jabberwacky incorporates AI and will keep track of the conversation that all users have submitted to it. It then tries to reuse those statements by matching them to the user’s input. Jabberwacky has no long-term memory, so it will respond only to the last sentence written. In future responses, Jabberwacky uses the learnt material from the recorded conversation logs. In some ways it models the approach that humans learn about language, facts, context and rules [85].

Nonetheless, these simple gambits can produce surprisingly intelligent-seeming conversations. That’s because they rely on a trick of human psychology making the systems to appear as more intelligent than it actually is. It has been noted that some

users have chatted with ALICE and Jabberwacky online for hours, apparently not knowing - or perhaps not caring - that they are conversing with machines.

But could one CA fool another CA? What would one say to another in private? To find out the result of the exchange, a conversation between ALICE and Jabberwacky was arranged and recorded<sup>35</sup>. To get each snippet of chat rolling, a “seed” question is posted from one CA to the other. After that, they were left on their own. What follows is a part of the unaltered record of what has been uttered by the two machines as shown in Figure 2.9.

*(Given seed question: Do you know the meaning of life? )*

ALICE	: The meaning of life is part of God's mysterious plans.
Jabberwacky	: <b>Do you know somebody that could know the meaning of life ?</b>
ALICE	: I don't know if I know it. Where would I have heard about it?
Jabberwacky	: <b>You always be here. These are the things that you have to know.</b>
ALICE	: Which ones?
Jabberwacky	: <b>Languages.</b>
ALICE	: LANGUAGES is a good topic. Sure, LANGUAGES sounds like a good topic.
Jabberwacky	: <b>Which language do you like?</b>
ALICE	: AIML.

**Figure 2.9: Example Conversation between ALICE and Jabberwacky**

When two strangers engage in a conversation for the first time, it is likely that they will try to initiate topics on common life experiences or common interests. Examples are the language they used, what they do, where they live or where are they from. A CA does not have life experience nor it has any “personal interest”. What they have is some canned answers that they try to match to questions posed by the users. While most of the CAs have some logging facilities to record the conversation, they do not have

---

<sup>35</sup> The conversation between ALICE and Jabberwacky take place at <http://www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1> and <http://www.jabberwacky.com/> on 27 March 2008

learning capability nor do they utilise such information. In other words, it could be said that the CAs do not have “long term memory”. They “forget” what was said a moment ago and their responses completely disregard the current context.

### **2.4.3 Commercial Conversation Agents System**

With the growth of electronic commerce on the WWW, CAs are becoming very useful in the commercial world. Commercial CAs are designed to fulfill specific functions. For example, a CA could be a part of a web-based customer relation management system. A CA could behave as a tireless employee - working 24/7, offering a friendly and useful first contact to customers, and are able to deal with the most common problems or requests for information as included in their knowledge base. The sole objective for these CAs is to be good at their particular job, rather than to be an expert in general conversation. However, nothing prevents the CAs from being able to talk about a wide range of topics depending on the amount of knowledge implemented by their botmasters.

In this section, the use of CAs for commercial purposes is investigated. As economic interests are behind the commercial programs, results from such approaches are expected to be more in focused than in the experimental approaches of the previous CAs. Classical and Loebner Prize CAs are mostly experimental, and the documentation of these agents are widely available. Since no proper documentation is available on commercial CAs, our investigation on commercial CAs will be based on the task environment of ELIZA; on the Performance measure, Environment, Actuator, and Sensor model (PEAS) [86]. There are a number of commercial CAs available for e-commerce and e-services, such as Anna<sup>36</sup>, Spleak<sup>37</sup>, Lucy<sup>38</sup>, Cybelle<sup>39</sup>, Verbot<sup>40</sup>, among

---

<sup>36</sup> <http://www.ikea.com>

<sup>37</sup> <http://www.spleak.com>

others. However, only two of them, namely Anna and Spleak, will be discussed in detail based on their contributions, the employment of the up-to-date technology and reputation amongst their customers.

#### 2.4.3.1 Anna

Anna is a digital assistant developed by Artificial Solutions for IKEA using Kiwilogic Lingubot. Digital assistants such as Anna are programmes that appear on e-commerce Internet sites in a variety of roles; their use will continue to grow as more companies see their deployment as enhancing human-computer interaction while building brand awareness and increasing sales. ELBOT, Loebner's 2003 bronze runner up and Chatterbox 2003 winner, is the underlying technology behind the text-based dialogical query system Anna. As a 24-hour virtual customer service agent, Anna, through text-based dialogue, assists customers to find items and prices from IKEA's online catalogue [53]. Anna engages in 20,000 conversation exchanges per day across IKEA Internet pages in eight countries and six languages, including English<sup>41</sup>. Aside from the textual dialogues, the webpage with Anna also presents an animated image of Anna's talking head without speech synthesis. The system deploys a URL push technique by sending a "pop-up" related to customer's request. As Anna is a commercial product, little has been published about the techniques behind it. Therefore, we evaluated Anna based on its behaviour by seeding a conversation from their online website<sup>42</sup> as shown in Figure 2.10.

John : Hi, my name is John  
Anna : **Thank you very much, John. Nice to meet you.**  
Anna *smiles*  
John : I'm looking for a table.

---

<sup>38</sup> <http://www.speak2me.net>

<sup>39</sup> <http://www.agentland.com>

<sup>40</sup> <http://www.verbots.com>

<sup>41</sup> Kiwilogic. Virtual Service Agents <http://www.kiwilogic.com/>

<sup>42</sup> [http://193.108.42.79/cgi-bin2/IKEA\\_USA.cgi](http://193.108.42.79/cgi-bin2/IKEA_USA.cgi)



**Anna** : **On this page you will find Tables & desks.**  
*A page pops up with variety of tables.*  
**John** : I need a computer table  
**Anna** : **Here you will find the Computer workstations.**  
*A page pops up with computer workstations table.*

**Figure 2.10: Example Conversation with Anna**

From the conversation above, Anna is able to keep and remember an internal state that reveals information such as names, and the most recent lines of input are remembered. The simplest selection method is the one used by simple reflex agents such as ELIZA. This selection mechanism uses regular expressions for identification of input and possibly for information extraction using ELIZA's transformation rule, which can be expressed as:

```
((I'm (/NAME))  
(Thank you very much, NAME. Nice to meet you.))
```

The conversation shown in Figure 2.10 also showed that Anna is able to store the current state describing the topic as the input "table". This context recognition will be an answer to a question and an expected response mechanism as used in the AIML architecture by using "topic". Such a feature requires a very limited world of IKEA furniture model. Even though Anna has a commercial purpose, the applied techniques demonstrate nothing more advanced than what ELIZA did. As claimed by Fred Roberts [87], Kiwilogic Editor does not build AI, but rather commercial dialogue systems with a well-defined area of expertise; usually company FAQs and Website navigation. Among these user inputs will be found commonly asked questions, which the Anna has not yet been prepared to answer.

### 2.4.3.2 Spleak

Spleak<sup>43</sup> is a virtual person created by Morten Lund at IMT Labs for instant message dialogues via Microsoft MSN Messenger. Spleak won second prize in Microsoft's Robot Invasion Competition in 2006<sup>44</sup>. As a virtual person, Spleak has the ability to participate in textual dialogues and appears in the form of a 21 year old girl from New York. It has the ability to chat with people and has access to data such as encyclopedias, dictionaries, weather forecasts, news, and entertainment items such as horoscopes and textual games, black jack, and hangman. Spleak also features a set of tools such as a calculator, unit conversion, spell checking, and web searching.

Similarly to Anna, Spleak is a commercial product (subsidiary of the Microsoft Corporation) and insight into the techniques used is thus not available. Therefore, we evaluated Spleak's behaviour by chatting with her, after adding her account ([spleak@hotmail.com](mailto:spleak@hotmail.com)) into our MSN Messenger contact list as shown in Figure 2.11. Spleak uses an information extraction technique to gather information about the human participant and is able to communicate with limited use of emoticons. Information such as name, age, and location is stored in a simple internal model for later use. However, a simple repetitive test will show if a model of the discourse is kept. When we repeatedly asked the same question, Spleak actually noticed the loop:

```
John      : are you there?
Spleak   : I'm here!
John      : are you there?
Spleak   : Yup, ready as I'll ever be.
John      : are you there?
Spleak   : One more time, 0!!! I didn't get it the first three times... 😊
```

**Figure 2.11: Example Conversation with Spleak**

---

<sup>43</sup> <http://www.spleak.com>

<sup>44</sup> <http://www.robotinvaders.com>

If two questions are asked repetitively in an alternating sequence, Spleak will not notice. Thus Spleak does not keep a model of the whole discourse, but at most remembers the last three inputs. Spleak also uses an aggressive strategy to abandon questions which are not understood and extract information from the human participant by dominating the conversation by asking questions. The intention of the topic shift is to avoid talking more about the user's interest. This mechanism is used in ELIZA by eliminating words from the input and AIML in their random <li> tag. This indicates that responses are chosen either on the basis of matched regular expression, or via a solely random mechanism. It is clear that a grammatical analysis of the input is not deployed.

Here, two commercial CAs have been investigated: IKEA's virtual assistant Anna and IMT Labs' virtual person Spleak. Although no documentation is available for either of these CAs, their behaviour indicates that a reflex agent design with a very limited model is used. The model is only capable of remembering simple information such as the name of the human participant in the dialogue. Unfortunately the commercial approaches seem to be nothing but direct imitations of ELIZA or ALICE, which could be the result of a limited budget or tricks, as the following section outlines.

#### **2.4.4 Tricks or AI**

Fifty five years after Turing proposed the Imitation Game, Weizenbaum's ELIZA program demonstrated that "a simple computer program" could successfully play the Imitation Game by resorting to a few "tricks", the most obvious being to answer questions with questions [32]. This also quite evident from the Loebner competition [23], the popularity of CAs based on AIML language [41], and the general lack of progress in text understanding and natural language dialogue systems. Cheating

obviously has its limitations, and it is doubtful that good natural language interfaces may be built this way.

Some people interpret the TT as a setting in which you can "cheat". The game has no rules constraining the design of the machines. The TT in general and the Loebner prize in particular reward tricks [88], and the winning programs for the last fifteen years have clearly incorporated some tricks. This sort of qualitative assessment to program knowledge is exactly what the ELIZA is trying to avoid, replacing the question "Can machines think?" with a performance test. Turing's Imitation Game is generally inadequate as a test of intelligence, as it relies solely on the ability to fool people. Perhaps the biggest obstacle to the advancing in this area is the fact that there are not many uses for fooling people besides the TT [89]. This can be achieved easily, as Weizenbaum has found [90]. Here we describe and compare some of the better tricks, confident in the belief that when a computer program eventually does pass the TT, it will use many of them. There is also a simple reason for this as people are already using them in everyday life as depicted in the Table 2.4

**Table 2.4: Conversation Agents' Tricks**

<b>ELIZA</b>	<ul style="list-style-type: none"> <li>- Fostered by including substrings of the user's input in the program's output. User: You hate me. ELIZA: Does it please you to believe that I hate you?</li> <li>- Use of the Rogerian mode, which provides unimpeachable cover for the computer.</li> </ul>
<b>PARRY</b>	<ul style="list-style-type: none"> <li>- Admitting ignorance. <i>"I don't know"</i>.</li> <li>- Changing the level of the conversation <i>"Why do you ask that?"</i></li> <li>- Introducing new topic, launching into a new story also called as simulates paranoid behaviour.</li> <li>- Launching into a new story.</li> </ul>

---

<b>OTHER TRICKS</b>	<ul style="list-style-type: none"> <li>- Having many fragments of directed conversation stored in activation network.</li> <li>- Using humorous statements to make the program seem more human.</li> <li>- Agreeing with the user, in certain cases, the program can safely agree with the user rather than being non-committal.</li> <li>- Simulated typing, by including realistic delays between characters that imitate the rhythm of a person typing.</li> <li>- Make longer replies that seem more human-like than current reply.</li> <li>- Introduce a new subject to divert the human user's attention.</li> <li>- Fragments of directed conversation stored in activation network, "<i>opening book</i>".</li> <li>- Controversial statements, eg "<i>People don't own cats...</i>"</li> <li>- Agreeing with the user, rather than being non-committal.</li> <li>- Excerpting News, weather forecast, textual games, etc.</li> </ul>
---------------------	--

---

## 2.5 New Challenges

In this thesis, the challenge for the development of intelligent CAs will not only be based on the central role of natural language, but also on the capability of the CAs to gather knowledge, and to act and speak like a human which will be explained in the following sections.

### 2.5.1 Natural Language Understanding (NLU)

The NLU field took off during the 1970s, beginning with Terry Winograd's SHRDLU [91] [92] system, a classical CA embedded in a world of toy blocks at the MIT Artificial Intelligence Laboratory in the late 1960s. SHRDLU focused on understanding natural language and AI, rather than responding in a human-like fashion. The domain was limited to a simulated world, which contained a table with a box and a number of coloured toy objects. The field of NLU is strongly connected to the field of AI. The aim is to transform the text into something that computers can "understand". That means that the computer should be able to answer natural language questions about the text, and also be able to reason about facts in different texts.

NLU paradigms use predicate logic as a semantic representation, such as that used by the question answering systems LUNAR[93], AskJeeves[94], AnswerBus [95] and

START[96], which is very important when it comes to choosing the best approach to building useful systems. These challenges highlight the strengths and weaknesses of the different systems for building further NLU systems. New challenges are being considered, including metrics for conversation naturalness and “meaningful” interaction with the users.

### 2.5.2 World Knowledge

NLU is probably not something that can be achieved merely on the basis of linguistic knowledge, such as knowledge of a grammar and lexicon or ontology. It probably requires much world knowledge - not only semantic and pragmatic knowledge, but also something like what AI researcher Lenet [97] calls “commonsense knowledge”; i.e. lots of facts about the world, and perhaps also some specialized domain knowledge (e.g. knowledge of how to play chess). The computational linguist Terry Winograd [92], who pioneered the use of NLU for CA development, argued for the claim that NLU requires such knowledge.

Three popular common sense knowledge bases have been created this century, namely Cyc<sup>45</sup> [98], OpenMind<sup>46</sup> [99] and Mindpixel<sup>47</sup> [44]. The OpenMind Common Sense project differed from Cyc because it focused on representing the common sense knowledge it collected as English sentences, rather than using a formal logical structure. This means that the information didn't use a standardized vocabulary with strict definitions for each component of the common sense knowledge. In addition, the Cyc model and OpenMind had a drawback which prevented truly large-scale collaboration [100]. Mindpixel, another AI commonsense project, was created using a web-based collaborative[100] AI project. MindPixel creates its knowledge bases by accepting input

---

<sup>45</sup><http://www.cyc.com>

<sup>46</sup><http://www.openmind.org>

<sup>47</sup><http://www.mindpixel.com>

from the general public, rather than knowledge experts. Some MindPixel data (GAC-80K) is being utilised by Cornell University [100] and The University of Memphis [101] to study theories in high-level reasoning. Some examples are the development of the Jackie [44] and ALICE Silver Edition [102] to simulate human conversation. Therefore a large-scale public collaboration collecting commonsense knowledge will play an important role in the development of the CA's knowledge bases.

### **2.5.3 Human-machine Interface**

In the search for the next generation user interface, there has been a growing interest in the development of an interface for CAs not only on the web, but also on instant messaging and mobile services. During the past decade, one can observe that there are rapid advances in CAs, spoken language technology and multimodal interfaces [103], [104], [105], [106] and [107]. All these have stimulated interest in a new class of conversational interfaces. As observed in the literature, researchers in this discipline have proposed different techniques and produced several natural language conversation systems.

Moreover, CAs represents the convergence of new interface agents, spoken language and human-computer dialogue systems. As CAs become more realistic, users expect to be able to interact with the agents in natural language.

## **2.6 Summary**

This chapter has given an overview of CA developments and human-machine conversation has been described. However, the Turing Test remains an issue in many discussions on AI. The Loebner Competition has been held every year since 1991.

Based on the transcripts<sup>48</sup> of these events, it is apparent that in terms of the techniques not much has changed, improved, or progressed [18, 108, 109].

One additional problem with TT, besides its reliance on a limited definition of intelligence, is related to natural language processing. The problem is exponentially more complex when one tries to develop a computer program to converse with a human. Even award-winning CAs including ALICE, which has been named the most indistinguishable from a human being in the Loebner Prize competition, have quirks that reveal them as not being human within the first five or six sentences of the conversation [38]. Jabberwacky comes a little closer to human language since it learns how to converse from previous user interactions. Anna and Spleak, two commercial CAs serve well as business identities. Based on the dialogs from these CAs, it appears that their responses are not much advanced or human-like than those from ELIZA and ALICE. Certainly, understanding unbounded natural language is one of the challenges within this study but it is not the sole research focus. The key contribution of this thesis has been the development of the CA framework offering interoperability, scalability and modularity, and the evaluation of such framework. However, there are several fine-grained conversation knowledge bases such as AAA and Mindpixel which make the processing of natural language tractable. These knowledge bases have been incorporated in the proposed framework. These facilities the development of the framework and focus on the practical aspects such as machine understanding, natural man-machine interface and equipping the CA with world knowledge as detailed in Chapter 3.

One of the features in proposed man-machine interface is the CA's ability to communicate via the Web, mobile services and instant messaging (IM) systems. These

---

<sup>48</sup> <http://www.loebner.net/Prize/loebner-prize.html>



aim to improve the human-machine interfaces with lifelike avatars and emotional expressions as an embellishment of the system. As shown in Table 2.2, none of the CAs previously developed are embedded with lifelike avatars or speech synthesis, except for the ALICE, Jabberwacky and Anna CAs. These CAs have deployed an animated ‘talking head’, but without speech ability. While some users may be content with a ‘text only’ interface, other users may be more comfortable with an animated avatar that provides visuals and sound. In the proposed system in this study, plug-in modules can be used to enhance the interface of a CA based on a flexible and modular approach as described in the next chapter.

## CHAPTER 3

### CONVERSATION AGENTS FRAMEWORK DESIGN

#### 3.1 Introduction

In this chapter, the challenges of developing reusable, extensible, scalable and modular software for CA frameworks are summarised. Solutions for these challenges are illustrated and discussed in the development of the CA called Artificial Intelligent Natural-language Identity (AINI). Most of the other CA designs are based on a two-tiered or three-tiered approach. The main contribution of this chapter is the proposed AINI framework with modular design which is based on a modified *N-tiered* architecture similar to the current Service-Oriented Architecture (SOA). Two additional tiers are proposed in the AINI framework: the *channel service tier* and the *domain service tier*. They can be implemented in a number of ways to provide the flexibility and scalability for different application domains and tasks. This has been demonstrated by two specific application domains on SARS [110, 111] and Bird flu [112, 113]. This chapter describes and exemplifies the underlying technologies being used in this thesis.

#### 3.2 Conversation Agents Framework

Developing reusable CA software is particularly difficult because there is no universally agreed definition of what a CA is. In addition to the proprietary commercial CAs, most of those developed are built on an *ad-hoc* basis with an aim just to pass the Turing Test (TT) or for entry in the Loebner Prize. Over the past fifty years since the birth of AI, CAs has taken many forms. Some of the CA researchers were aspired to the development of CA embodiment known as embodied conversation agents (ECAs) [103],

[104], [105], [106]. Other CAs were designed for particular domains such as edutainment, medical, customer relationship management (CRM) [67, 69, 114-116], or crisis communication applications [117] [82]. On the human-computer interface (HCI) aspect, these CAs took the format of Computer Mediated Communication (CMC) using instant messages, e-mails and chat rooms [118-123]. The communication channel and technologies could be both wired and mobile services including SMS, MMS, WAP, or GPRS [124-126]. In terms of expression, these CAs may use verbal (speech synthesis, voice recognition), non-verbal (i.e., text-based, emoticons) [127, 128], and even reconfigurable animated characters [10, 59, 129, 130] with different faces, hands, heads, clothes, genders or other accessories.

It is obvious that the design of modern CAs requires a better structure for their implementation. The goal is to develop a practical framework to improve the interoperability of the CA architecture through modular design. This is illustrated in the AINI framework [112, 113, 117]. Finally, the proposal is to build the CA based on established and stable multi agent framework. This reduces the needs to rewrite the complete application. These attributes can be summarised as follows:

### **3.2.1 Reusability**

Framework reusability implies the reuse of the domain knowledge, interface design and query engine in order to avoid recreating and revalidating common solutions to recurring application requirements and software design challenges. Developing reusable frameworks could be a tough job and involves much effort, but it is generally believed that the work pays off at the end. Reuse of framework design components can yield substantial improvements in productivity, as well as enhancing the quality, performance, reliability, and interoperability of software.

### 3.2.2 Modularity

In the AINI CA design, this framework enhances modularity by encapsulating implementation details and helps improve software quality by localising the impact of design and implementation changes. Creating modular architectures will allow plug-and-play or plug-in modules to reconfigure software variations to meet a specific design.

### 3.2.3 Extensibility

Extensibility in AINI framework design is essential to ensure timely customisation of new application services and features. A framework enhances extensibility by providing the means to allow applications to extend its interfaces and functionalities. These extension points or service modules systematically decouple the interfaces and behaviour of the application domain. For instance, AINI uses XML specification because the standard is already gaining widespread support. This has been generally accepted by the community as the preferred choice of knowledge and messaging representation language.

### 3.2.4 Scalability

In the AINI design, the framework must provide scalability features where the design can be modified without necessarily affecting other modules or the rest of the architecture. This relates to:

- *Conversation module*

New methods should be able to add to the conversation module by generating new services to the behaviour modules in the CAs. This module supports web-based, mobile-based or other network-based communication protocols.

- *Number of conversations and participants*

The system should be able to accommodate an increasing number of participants and to keep track of multiple conversations without modification of the architecture.

In addition, the conversation logs should keep the records of the conversations of the participants. This information could be kept for future analysis or compliance requirement.

### **3.3 Conversation Agents' Features**

There are many challenges in the development of AINI, these include the design, issues related to the flexibility and extensibility of the system, maintaining simplicity without being simplistic, extendibility, and long-term maintainability. The real challenge is how to provide a framework in which a developer can work on the CA system without having to understand the all modules within system. This challenge is also to find the delicate balance among the above elements with regard to the following aspects.

#### **3.3.1 Modules Integration**

The most challenging element in developing conversation software is the integration of all modules by establishing a plug-and-play architecture. To keep the complexity of systems manageable, and to simplify the testing and maintenance of the various modules, it is important to reduce code duplication as much as possible across the domains.

#### **3.3.2 Domain Independent**

Another challenge is the inherently complex multi-domain nature of the AINI system. Developing a generic framework requires continuous refactoring of common elements across multiple domains. In addition, building world knowledge and domain-specific knowledge for each application and task are labour intensive; therefore automated tools are necessary to manage this process.

#### **3.3.3 Cross-Platform**

A challenge in developing CA software is to decide the appropriate computer software platform. This includes operating systems, browsers or programming languages design

[131] - CAs can be implemented on multiple computer platforms such as Windows, Linux or Mac OS. The computer software: LINUX, Apache, MySQL and Perl, known as “*LAMP*”, is the solution stack. These are free open source programs, representing a non-proprietary, flexible way to create a client-server based application. The combination of these technologies is used primarily to define a web server infrastructure.

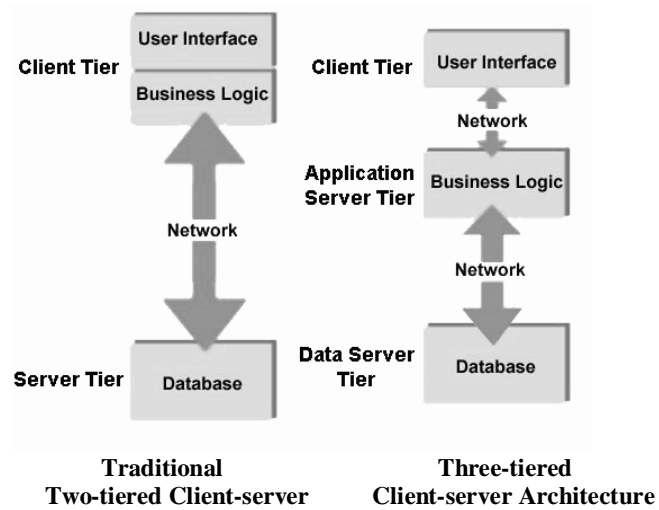
Common computer languages used to design classical CA are mainly based on Lisp, Prolog, Icon, Bash, C, Haskell and Curl, which do not rely on client/web-server architecture. Although the Java Program dB [132] won second place in the Loebner Prize, it has the disadvantage that it hard-codes a local file path in the AIML file. Program dB is based on two-tiered architecture: the targeting server and the Loebner contest text-based interface. Program dB has no "long term memory" and it forgets all the clients each time it is shut down. This is partly the result of the lack of database, but it seems the approach improves the overall performance of the server. In this study, the CA development was based on cross-platform programming and LAMP solution, which will work on multiple platforms. Our principle was that the language must be simple, small in size, speedy and powerful [133]. Perl definitely made imperative programming concepts and object concepts much easier to grasp and use. This is one of great features of Perl.

### **3.4 N-tiered Architecture**

In recent years there has been a dramatic growth in distributed computing environments. Unfortunately, the size and variety of heterogeneous computer systems make distributed computing difficult and complex [134]. To make systems interoperable, middleware is usually required. Middleware is software that rewraps the architectural details of a system and offers services to other systems on heterogeneous platforms across a network [135].

In the 1980s, the arrival of inexpensive network-connected personal computers produced the popular two-tiered client-server architecture[136]. In two-tiered architecture, every time the business rules are modified, the client application has to be changed, tested and redistributed, even when the user interface remains intact. Therefore, the two-tiered architecture suffers from scalability and modularity issues.

The third tier (application server tier) is between the user interface (client tier) and the data management (data server tier) components. This middle tier provides process management where business logic and rules are executed and to accommodate hundreds of users by providing functions such as queuing, application execution, and database staging. The three-tiered architecture is used when an effective distributed client/server design is needed to provide an improved performance, flexibility, maintainability, reusability, and scalability [137-139]. A comparison between two-tiered and three-tiered client-server architectures is shown in Figure 3.1.



**Figure 3.1: Comparison Two-tiered vs Three-tiered Client-Server Architecture**

The three-tiered system architecture is more commonly referred to as  $N$ -tiered architecture in reference to the unlimited number,  $N$  intermediary layers between the client and server, popularised by Bonasso[140].  $N$ -tiered applications are highly scalable and offer the best performance. They can support any combination of user interfaces: web browser, mobile interface, network interface and others while offering the highest potential for code reuse and sharing.

### 3.5. AINI's Conversation Agent Architecture

$N$ -tiered architecture has been deployed by numerous domain applications and agencies, such as Human Capital Management system (HCM)[141], DARWIN [142] designed by NASA, SLAM 3DSim mobile robot[143], Urban Search and Rescue[144], Carnegie Mellon Navigation (CARMEN) Toolkit [145], CoCo an autonomous agent[146], spoken dialogue [147], mobile service software agent Open APIs(Parlay/ JAIN) [148], storytelling robot [149], Learn Sesame [150] and Intelligent ChatBot [151] to name a few.



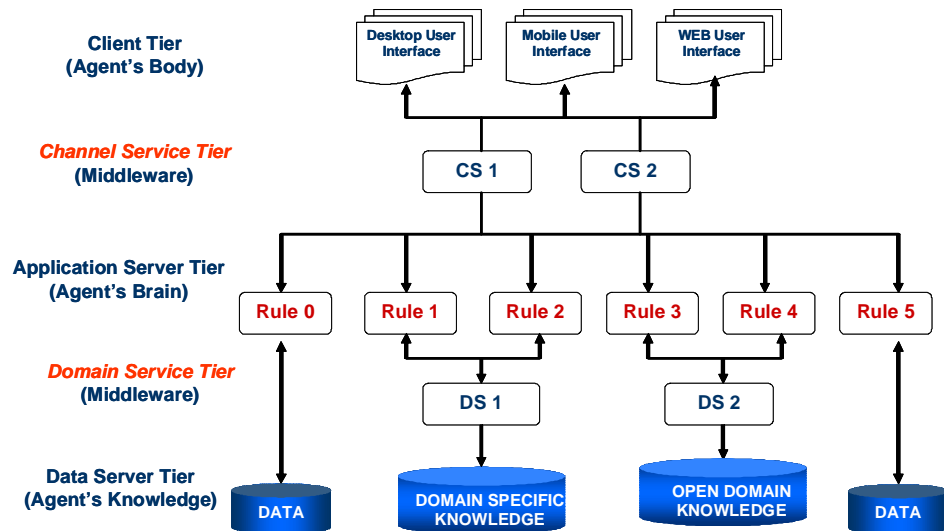
In the design of CAs such as TARA(Terrorism Activity Resource Application) [83, 84], CAs were experimented with using a two-tiered architecture based on the original ALICE's Loebner Prize winning Program D [152] on Java. Although Program D can be configured with unlimited number of CAs, their handwritten AIML knowledge bases are hardcode in XML and precompiled with Java, instead of using a flexible relational database [153]. This was followed by CMU Nursebot called Pearl [78, 79] and OpenMindBot [80], CatBot [64], TutorBot [65], Partner (SP) [66, 67], ChatBot [151] and Persona-AIML [68]. However, VPbot [74, 75], the Harvard Medical School's Virtual Patient program, goes beyond this limitation by deploying Program E [154]. Program E is based on PHP programming language, which supports multi-tier applications. VPBot makes use of a relational database by loading the AIML files into a MySQL database.

While many techniques and programming languages have been proposed over the years to develop CAs, the primary challenge remains in how to overcome the poor scalability and the lack of flexibility to handle the heterogeneous CAs software designs. Therefore, the proposed approach is to employ a novel modified *N*-tiered architecture similar to the *service-oriented application architecture* to capture information at decreasing levels of granularity at the agent brain (application server tier), agent body (client tier) and agent knowledge (data server tier). This architecture enables us to handle dynamic features unique to scripting languages, such as dynamic typing and code inclusion, which have not been adequately addressed by previous techniques.

### **3.5.1 AINI's Modified *N*-tiered Architecture**

The key to success in future conversation systems is to develop general dialogue systems [155]. This can be defined as a framework that is not designed for a particular

application, but can be ‘plugged’ to various applications. It should require minimal effort to develop the domain knowledge and to adapt to different applications. These considerations have led to the proposal of two main tiers for the handling of the generality issue in the CA framework. They are the *Channel Service Tier* and *Domain Service Tier* as shown in Figure 3.2.



**Figure 3.2: AINI's Modified N-tiered Architecture**

These additional tiers contain all the necessary logic to receive and delegate the requests. In the previous three-tiered architectural concept, it was illustrated that the architecture is lacking in two major aspects. First, the three-tiered architecture combines communication logic with business logic in the application server tier to handle all the requests by using a single interface to access services. Second, in the three-tiered architecture, clients and developers are faced with potential issues with the communication channels. In order to avoid these consequences, the *Channel Service Tier* has been used, as it allows the creation of a flexible and extensible architecture able to support different interfaces and additional services more easily. The *Domain Service Tier* will support a flexible and extensible architecture that is capable of providing any

domain application based on a client's request, without the necessity of changing domain services.

#### **3.5.1.1 Channel Service Tier**

Building extensible and scalable systems is a challenging task. Currently, different clients communicating over different channels or protocols such as HTTP or TCP, must be offered the same set of application services by the system. In addition, the current deficiency of CAs architecture is the lack of universally agreed standards. In general, the number of channels and the number of demanded services are expected to increase over time, and a gradual shift from web-based to mobile-based applications. The proposed AINI's modified *N*-tiered architecture could be a viable solution to meet these future challenges. Each channel has its own communication protocol. As the application tier has to satisfy requests from clients such as web browsers, mobile browsers, MSN Messenger interface applications, it must deal with service requests coming from different channels and each channel is characterised by its own communication protocol. Therefore, a server residing on the application tier must be able to support these communication protocols.

#### **3.5.1.2 Domain Service Tier**

Under the standard *N*-tiered architecture, the application tier and the data access tier are flexible and scalable toward domain knowledge independency. The suggestion made here is to create an additional layer called the "domain service tier" to resolve the coupling issues. The role of the domain service tier will be focused on handling concurrent access and managing changes to the domain knowledge tier. This domain service tier prepares a number of its own services that override methods or features of the database using SQL query. It is also the responsibility of the domain service tier to choose either the domain-specific or open-domain knowledge (discussed in section

3.5.3). This solution increases the cohesion of each tier and prevents the overall system from becoming excessively complex. In addition, it will assist the development of the domain application. For instance, to deploy a furniture domain application such as the Anna CA used by IKEA (see section 2.4.3 Commercial CAs), only domain-specific need to be updated, instead of the entire data server tier.

The AINI conversation architecture has been reported in previous publications [117, 156]. As illustrated in Figure 3.3, AINI employs a modified  $N$ -tiered architecture that can be configured to work with any web, mobile or other network applications. It comprises a client tier (hereafter called “*agent body*”), an application server tier (hereafter called “*agent brain*”) and a data server tier (hereafter called “*agent knowledge*”).

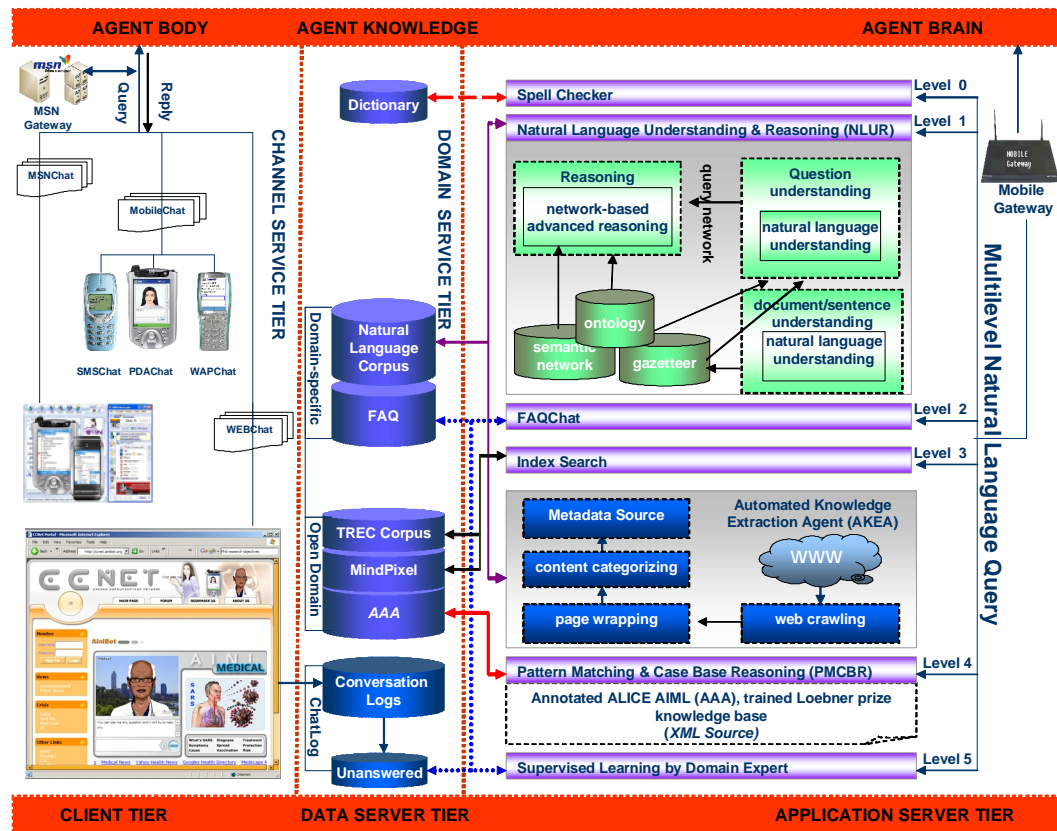


Figure 3.3: AINI's Conversation Agent Architecture

As mentioned previously, the *channel service tier* and the *domain service tier*, are designed to support different channels of interfaces and can quickly be augmented with domain knowledge for specific purposes. The architecture provides features of multimodal interface, multilevel natural language query and multiple knowledge bases. The process of communication and answering is as follows. Given a question, AINI's agent body will perform a HTTP over TCP request from a Web, mobile service or instant messaging service to the agent brain. Here, the agent brain will attempt to formulate a reply to the sentence via the Natural Language Understanding and Reasoning Module. This module will do the sentence parsing to produce grammatical categories and grammatical relationships by extracting pertinent information through the agent's knowledge, such as Noun Phrase and Verb Phrase. If this step is successfully parsed, the network-to-path reduction will be carried out by Network-based advanced reasoning. From the query network, the *question* is decomposed into sequences of words or phrases and they form the nodes of a tree. After obtaining the query network, the task of answering the question is reduced to discovering the presence of the query network in the whole semantic network. Then, the problem of discovering the *answer* has been reduced. The appropriateness of answer is achieved through the logic and network approach during answer discovery using a template-based approach. The template method matches the answer against question templates requested by the user. The system will then produce an answer in the agent body. Each of the agent tasks is described in further detail in the following section.

### **3.5.2 Agent Body (*Client Tier*)**

The user interface, or human-computer interface (HCI), resides in the agent body and it supports three different types of channels of communication, such as Webchat, MobileChat and MSNChat, controlled by the *channel service tier*. AINI uses HTTP

over TCP to connect to the Internet and mobile services to communicate with the users. The user interface was written using HTML, Javascript, vbscript, XML, WML, Flash Action Script and other client side scripting languages. For the MSNChat, AINI connected to the MSN Messenger client through the MSN Mobile Gateway as shown in Figure 3.3.

### **3.5.2.1 WebChat**

WebChat is web-based interface that allows users to interact in real-time with AINI through cross-browser including Internet Explorer, Mozilla Firefox, Safari, Opera, and Netscape. For a few years, the CAs [104, 157, 158] emphasised the ‘embodiment’ feature of the interactive virtual characters or web robots [159] significant progress was made in terms of software architectures [160] [161]. To foster a relationship with the user and to encourage the user to interact with the system in a natural manner, it is possible to incorporate an anthropomorphic agent or avatar in the AINI’s framework. Nass and Reeves [162] from Stanford University suggested strongly that there is a tendency to treat computers as a human entity, which in turn results in human based social responses. Cassell et al. [158] pointed out that conversational computer agents have played very important roles in human–computer interactions. Such skills include the abilities to use face, hands and tone of voice to regulate the process of conversation, as well as the ability to use verbal and non-verbal means. To achieve this end, Microsoft Agent Technology [163], Flash Technology [164] and Oddcast Technology<sup>49</sup> [165] are used. These technologies introduce the use of an animated and speech enabled avatar, which will serve as the medium between the system and the user. Besides, users also can go through all the information on the website for the topics they are interested in. At

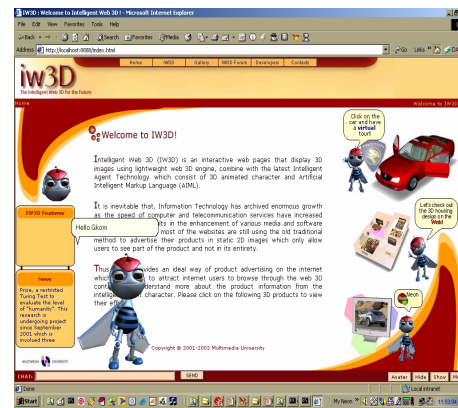
---

<sup>49</sup> Crisis communication research on SARS was supported by Oddcast Inc. in 2003 and Bird Flu in 2005. In these projects, Oddcast provide a streaming voice through text-to-speech (TTS) solution to make our ECAs more presentable. Oddcast interface that allows us to create and embed customize animated characters within AINI’s Framework.

the same time, they can place questions to AINI for more information or guidance. Another advantage of the WebChat is a collaborative browser which allows a portal to guide the users through the website of the organization by automatically “pushing” URLs and information from other websites to the user’s browser. This not only facilitates communication between CAs and users, but also allows the intelligent CA to help users locate specific information on their websites. In addition, users are also able to personalise their WebChat interface by customising the avatars, in such ways as choosing their avatar’s gender, voice, face, clothing or accessories as shown in Figure 3.4. As described in section 3.2.4, AINI’s architecture is scalable and can be embedded into any existing website such as a banking portal or education portal, and can even be integrated into a chatroom without changing its code, as shown in the Figure 3.5.



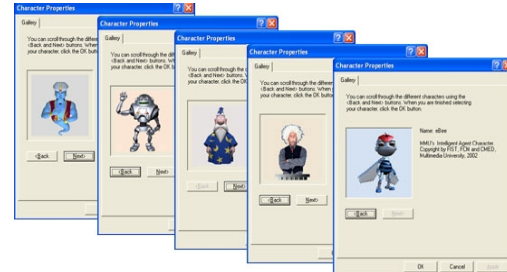
CCNet Crisis Communication website



IW3D – Intelligent Web 3D website



Choose your character, hair style, clothes, classes, etc. with Flash Technology



Choose your favourite 3D avatars with Microsoft Agent Technology

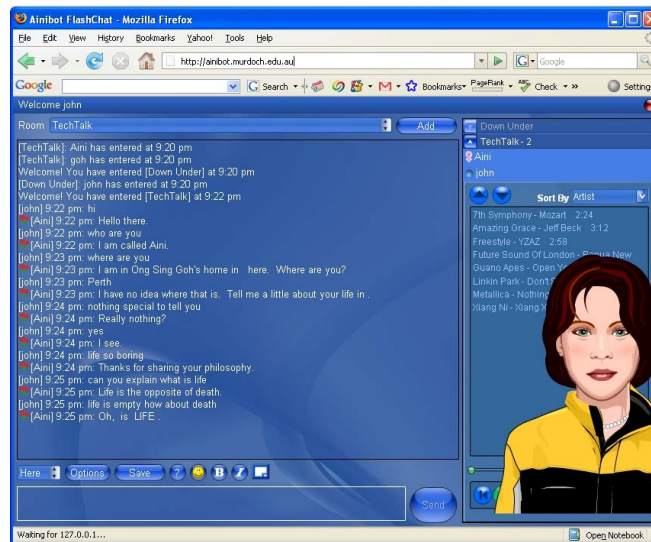
**Figure 3.4: Personalise User interface**



An illustration on how AINI could be embedded in a Malaysian Banking Portal



An illustration on how AINI could be embedded in the Murdoch University website with an Avatar



An illustration of AINI being integrated into the web-based Chatroom interface

**Figure 3.5: Examples to illustrate AINI's Scalable Interface**

### 3.5.2.2 MobileChat

Conversation chat through messaging applications were the first and most successful community applications for mobile services such as SMS, WAP, GPRS and 3G extended by Web services. In AINI's framework, users can freely select the CA they prefer to access. Hence, they can chat anywhere, at any time, with any device. A mobile chatting module is implemented in a series of logical phases. Based on the Research and Markets report [124], the text based agent-to-mobile chats with agent-to-Internet and Internet-to-mobile chats are likely to be popular and implemented in the future. Today, around one trillion text messages are sent each year, and this number is



growing. Here, AINI's modified *N*-tiered architecture is ready to support this technology as described in section 3.5.1.1. AINI's MobileChat module consists of SMSChat, WAPChat and PDACHat sub-modules as discussed below:

- **SMSChat and WAPChat**

The mobile chat is an alternative method by which users can chat with AINI using SMS, MMS, GPRS or 3G services. The SMSChat services are the text-based chatting system. MMS is a service similar to SMS but with added image, voice, animation and other features. Meanwhile, the WAP technology provides mobile web browsing functionality for accessing news and other forms of data services by connecting the WAP gateway with a given URL. WAPChat and SMSChat provide text-based interactive information services and applications from the screens of their mobile phones as shown in Figure 3.6 and Figure 3.7 respectively.



Step-by-step SMSChat with AINI

**Figure 3.6: SMSChat Interface**



**Figure 3.7: WAPChat Interface**



**Figure 3.8: PDACHat Interface**

- **PDAChat**

The idea of developing AINI into Personal Digital Assistance (PDA) is an interesting approach to having a more human and personalised interface between a computer and human [166-168]. The PDAChat with AINI performs functions similar to web chats, but in a mobile environment. It is a prototype designed to blend mobile technology with natural language to help humans interact more naturally with mobile devices. An example of PDAChat is shown in Figure 3.8. The PDAChat was designed using WiFi technology and powered by Microsoft Windows Mobile Technology [169] embedded with Pocket Internet Explorer on a HP iPAQ Pocket PC. The computer-generated avatar of the CAs is displayed on the PDA using Adobe Flash for Pocket PC<sup>50</sup>, and the system uses a WiFi wireless connection to AINI's server.

### 3.5.2.3 MSNChat

The flexibility and scalability of AINI's *Channel Service tier* not only allows web or mobile services, but also CMC application such as MSN messenger. MSN Messenger is a freeware instant messaging (IM) client that was developed and distributed by Microsoft<sup>51</sup>. The "language" used in IM communication between two computer programs is called a "protocol". The rules for messages sent between MSN Messenger clients' servers are called the "MSN Messenger protocol". The architecture of MSN Messenger is very complicated compared to other IM services such as AIM and Yahoo!, since it relies on five different types of servers to handle the communication and operation of its service<sup>52</sup>. MSN Messenger uses the Mobile Status Notification Protocol (MSNP) for communication. As shown in Figure 3.3, AINI's agent body uses MSN

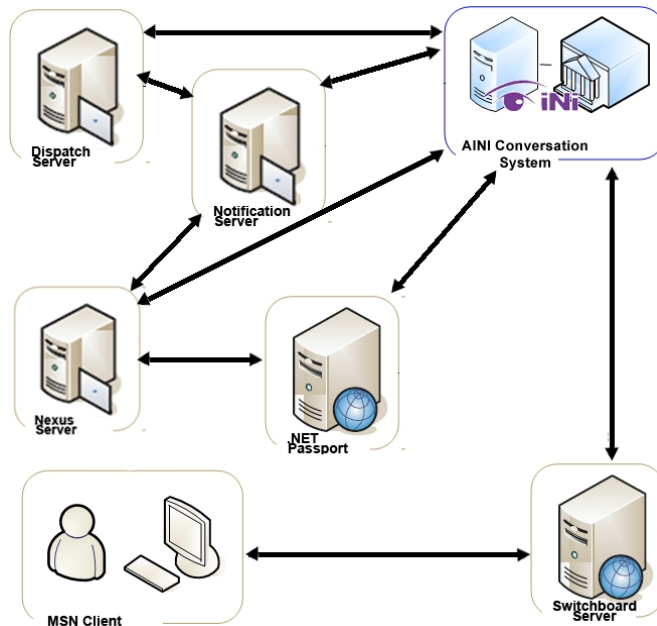
---

<sup>50</sup><http://www.adobe.com/software/flashplayer/pocketpc>

<sup>51</sup><http://get.live.com/messenger/>

<sup>52</sup><http://hypothetic.org/docs/msn/>

protocol to communicate with MSN Messenger servers. AINI utilises the .NET Passport to sign into the MSN Messenger service. MSN Messenger sign-in session is based on a challenge-response mechanism to authenticate user credentials. The communication with the Passport server is conducted over the HTTPS (Hypertext Transfer Protocol over Secure Sockets Layer) protocol, ensuring that the sign-in information is encrypted. The client sends the challenge string, Passport username, and password to the Passport URL. If the credentials for signing in are confirmed, the Passport server issues a ticket, which is passed back to the notification server to complete the authentication procedure. Figure 3.9 details the entire authentication procedure for AINI and MSN Messenger.



**Figure 3.9: AINI and MSN Authentication Process**

In the MSNChat module, we have outlined the conceptual and practical basis for the development of the AINI for MSNDesktopChat, MSNWebChat and MSNMobileChat sub-modules. All these modules are supported by the MSN Messenger protocol.

- **MSNDesktopChat**

MSN Messenger for Desktop, or MSNDesktopChat is the most popular free instant messaging client software [170] which captured 61 percent of the worldwide IM market share. MSN Messenger was developed and distributed by Microsoft Windows since 1999, and was renamed Windows Live Messenger in 2006. The major use of the software is IM, although other features which now come as standard include support for voice conversations, webcams, transferring files, and built-in two-player online games. Similar to many MSN Messenger's competitors, MSNDesktopChat allows messages to be enhanced with graphical emoticons (sometimes called 'smileys'), Flash animations called 'winks', animated display pictures, styled text, and many more accoutrements; with third-party add-ons as shown in Figure 3.10.



**Figure 3.10: MSNDesktopChat Interface**



**Figure 3.11: MSNWebChat Interface**

- **MSNWebChat**

The MSNWebChat module allows the users to interact in real time with the AINI software agent via a browser through MSN Web Messenger. It is possible for virtually any computer with an Internet connection to connect to the Messenger Service by using MSN Web Messenger. The MSNWebChat interface however does not have an interface

that mimics the MSN Messenger application as the MSNDesktopChat. Currently, these web-based chat sessions only support plain text, as shown in Figure 3.11. However, this MSNWebChat is a collaborative browser which allows a portal to guide the users through the website of the organisation. This is done by automatically “pushing” URLs and information from other websites to the user’s browser. This not only facilitates communication between the CA and users, but also allows the intelligent CA to help users locate specific information on their websites.

- **MSNMobileChat**

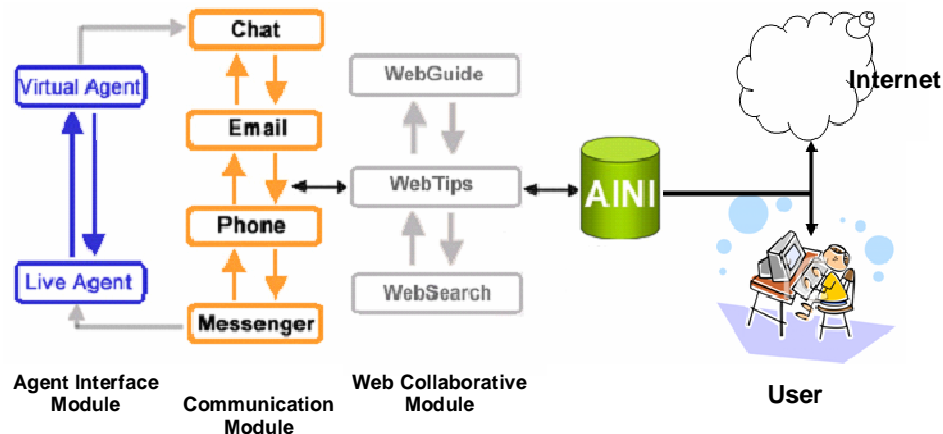
MSNMobileChat offers the same features as MSN Messenger for MSNDesktopChat. For MSNMobileChat, MSN Mobile device users can page AINI’s users with the `PAG` command and receive incoming messages with the `IPG` command. Mobile chat is an alternative way in which users can chat with AINI using GPRS, WiFi and 3G services. At the same time, wireless telephony networks using 3G offer smartphone users a subset of TCP/IP networking services such as IM. This service offers far broader coverage than WiFi technology and is powered by Microsoft Windows Mobile Technology [169]. In this interface, AINI provides text-based interactive information services and applications from the screens of their PDA or smartphone as shown in Figure 3.12.



**Figure 3.12: MSNMobileChat Interface**

### 3.5.2.4 Proxy Conversation Example 1

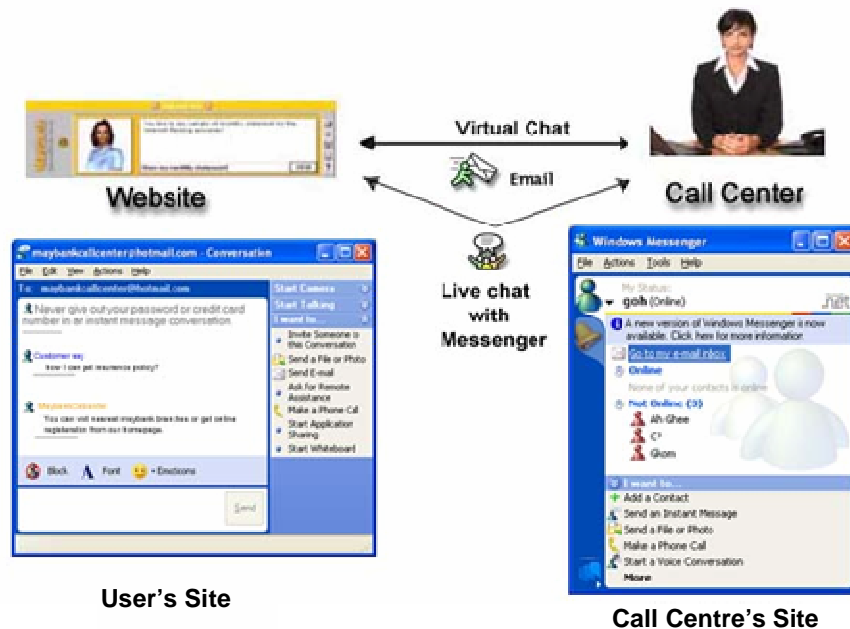
The agent body provides the necessary interaction between the user and the AINI control by using the communication channel service tier. In this proxy conversation<sup>53</sup> example, users interact with AINI through the normal Internet ports, which are connected to an agent body that provides communication between AINI and users with a Web Collaborative Module including WebGuide, WebTips and WebSearch as shown in Figure 3.13. This communication channel module has been developed in previous research [171]. The purpose of WebGuide is to guide users through the entire portal. It enables AINI to offer help without waiting for the user to ask. The WebTips engine will provide tips or hints to users. It is an intuitive feature that will recommend links within the site, whereas the purpose of WebSearch is to assist the users through the Google search engine. This web collaborative module functionality also allows AINI to “guide” the users through the organization’s Web site by automatically “pushing” URLs to the user’s browser (this function is also known as the “URL Push Technique”). This not only facilitates communication between AINI and the user, but also allows AINI to help users locate specific information on their Web site. In this way, the web collaborative module enables AINI to offer help without waiting for the user to ask.



**Figure 3.13: AINI's Communication Channel Module**

<sup>53</sup> This is a localhost proxy conversation which can be used with almost any client in order to participate in Off-the-Record conversations.

However, if certain predetermined conditions are not met and AINI is unable to retrieve the information required by users, then AINI will, through the other communication module, automatically forward an email form that can be filled and forwarded to a call centre. But if users require instant information, AINI will forward the telephone number of the relevant person in the call centre. This means that the users can call a live agent or human directly by telephone for further details. On the other hand, if the users feel that it is expensive to make phone calls (especially long distance calls), then AINI will offer the instant messenger service using MSN Messenger through agent interface module as discussed in section 3.5.2.3. The instant messenger enables the user to interact with live agents in the call centre via IM. The live agent at the call centre will be able to trace the discussion, as well as the user's profile as shown in Figure 3.14.



**Figure 3.14: An example illustrating a user communicates with Live Agent (human) through MSN Messenger**

The excerpts in Figure 3.15 are taken from a typical single session of proxy conversation exchanged between AINI and a user called “Ahmad” at the Maybank

portal [171]. This session illustrates the nature of human-machine communication. The conversation continues from the moment when AINI transferred the query via text-chat and voice-chat with the avatar. The chat-based system will allow the user to interact with the system in a manner that simulates a normal conversation. Some of the typical inputs include greetings, questions, answers, orders, remarks and comments, to which the agent will be able to respond.

#	ID	Message
1	AINI	Hi there :)
2	Ahmad	what's ur name?
3	AINI	My name is Aini. What is your name?
4	Ahmad	Ahmad
5	AINI	How are you, Ahmad?
6	Ahmad	Wonderful!
7	AINI	Good to hear that! I feel wonderful too!
8	Ahmad	Oh yes it is. Are you alive?
9	AINI	Yeah that's right. Our thoughts are one. Yes I sure am alive. It's great to be alive, isn't it?"
10	Ahmad	What exactly do you do?
11	AINI	I am a virtual tour guide for Maybank - The Largest Commercial Bank in Malaysia. I help surfers to get to know Maybank better.
12	Ahmad	Interesting. Tell me more about Maybank.
13	AINI	Apart from being the largest commercial bank in Malaysia, Maybank is also one of the most trusted bank by fellow Malaysians. We have an excellent customer service. We also provide a wide range of services to our clients.
14	Ahmad	That's so interesting.
15	AINI	Thank you for your humble compliment. You make me feel happy.
16	Ahmad	How am I supposed to open a new savings account?
17	AINI	First you need to choose our various type of savings account. There are Yippie Savings Account for young clients and normal Savings Account for everyone! If you wish to know more detail, I can transfer you there!
18	Ahmad	Thank you. I want to know more about Yippie Savings Account.
19	AINI	You are quite welcome. Yippie Savings Account is for young clients below 18 years old. The young clients will experience a lot of benefit from a holder of Yippie Account.
20	Ahmad	How can I know more about it?
21	AINI	I can transfer you there! Do you wish me to bring you to the Yippie account page?
22	Ahmad	Yes
23	AINI	OK. There you go ... <AINI uses URL "push" technique to forward a Yippie Account web page to Ahmad>
24	Ahmad	Thank you. Bye.
25	AINI	You are quite welcome. See you later.

**Figure 3.15: An example of Proxy Conversation from a Financial Portal based on WebChat**



### 3.5.3. Agent Knowledge (*Data Server Tier*)

Another significant difference between this research and other research on CAs is the domain knowledge model. Dahlbäck and Jönsson [172] stressed that the domain model represents the structure of the knowledge which comprises a subset of general knowledge. Such systems normally are comprised with two subcategories: the *traditional/narrow domain* or *domain-specific*, and the *open-domain*. In the traditional domain, systems attempt conversational fluency based on limited domains of expertise. ELIZA [35], for example, simulates a Rogerian psychotherapist, and its implementation is commonly known as DOCTOR and PARRY[33]. DOCTOR and PARRY's domain was restricted to paranoid hospital patient expressions. SHRDLU [92] is another program simulating a CA which is able to interact within a simple world knowledge of "blocks" (see detailed in section 2.4.1). SHRDLU was an entry in an early Loebner Prize competition, where the evaluation was based on the restricted tasks [109]. However, in the Fifth Annual Loebner Prize Contest in 1995, the Loebner prize criteria were changed to include unrestricted domains [27], requiring computer entries to converse indefinitely with no topic restrictions.

Hence, it is understood that general purpose CAs are not necessarily able to answer questions on a specific domain subject. On the other hand, domain-specific systems lack the flexibility to handle common sense questions. To overcome the above limitations, we proposed the Domain Knowledge Matrix Model (DKMM) [173]. The data server layer serves as storage for data and knowledge required by the system. This is where AINI's conversational knowledge bases are stored. It is well understood that true intelligent action requires large quantities of knowledge. Such a reservoir of knowledge can be harvested from the Internet and deployed in the domain matrix knowledge bases'

architecture. This forms the basis for the construction of large-scale knowledge bases to be used as the engine for intelligent conversation systems. AINI is the mechanism used to manage the knowledge and to provide appropriate answers to the user.

### 3.5.3.1 Domain Knowledge Matrix Model (DKMM)

AINI's DKMM incorporates several knowledge subjects. This is analogous to the consultation of expertise knowledge from multiple experts. For example, a *sales* knowledge domain should contain expertise on how to improve sales. However a sales person is expected to have a wide range of common sense, general knowledge or world foremost inventor's biographies [174] which enable CAs have ability to engage the potential customer in general conversation. Hence, an intelligent system should also incorporate open-domain knowledge to handle general or generic questions. By including multiple domain knowledge bases within AINI's single knowledge domain, the proposed AINI will be able to hold “meaningful” and prolonged the conversations with the users.

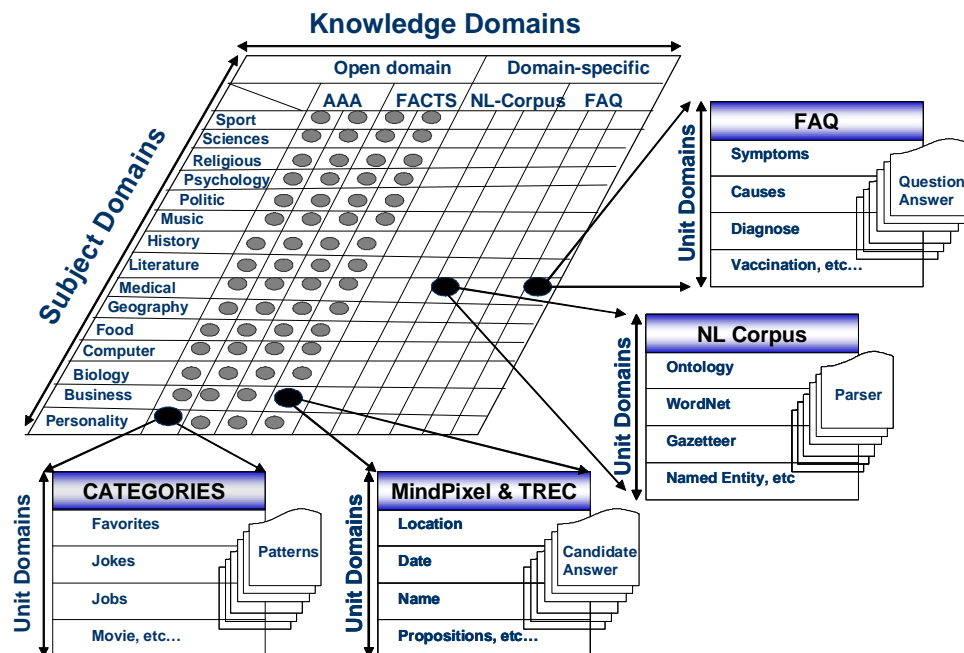


Figure 3.16: Domain Knowledge Matrix Model (DKMM)

In this proposed DKMM [175], both the open-domain and domain-specific knowledge bases are predefined in the agent's knowledge. These modules are used to support the various knowledge levels at the agent brain tier. Depending on the user's input, the agent will respond or switch from one level to another in the agent brain (discussed in section 3.5.4). While the system is capable to communicate with the user beyond the knowledge domain, there are cases where the system will exhaust its capability to answer the queries. In such case, the system will attempt to divert the focus back to the current topic of interest by responding with some predefined random statements. The purpose is to direct the user's attention back to the system's domain-specific state. Hence, AINI will attempt to "cycle" between the six levels of information processing within the agent brain tier supported by the various knowledge modules in the agent knowledge tier, discussed in Section 3.5.3.

A way to view the proposed DKMM is given in Figure 3.16. In this approach, the knowledge base of the AINI can be considered as a collection of specific conversation domain units. Each unit handles a specific body of knowledge used during the conversation between AINI and the user. The knowledge can be seen as arranged in the vertical columns making up the open-domain or domain-specific knowledge. In addition, specific subjects are shown in the horizontal rows. For example, in the open-domain knowledge, the subject units will cover topics such as personality, business, biology, computers, etc. In this research, our focus is on the subject of medicine; and in particular, the bird flu pandemic. Therefore, additional bird flu domain knowledge is being incorporated in the domain-specific row "medical", and column NL-Corpus (details on the extraction of the bird flu knowledge base from the web is discussed in Chapter 4).

In this research, the novel contribution is the development of the DKMM, which is a “*domain knowledge plug-in module*” through the domain service tier (see section 3.5.1.2). With this arrangement, the domain-specific knowledge and open-domain knowledge could become portable, scalable and incorporated easily with other domain applications. This approach will also allow future improvements to encourage collaborative contribution to the other domain applications and tasks.

### **3.5.3.2 Open-Domain Knowledge Bases**

Because we are merely making an approximation of the world’s general knowledge, any knowledge generated or manipulated in the domain of the system is collectively referred to as domain knowledge. Some literature refers this as ‘world knowledge’, where the use of the word “*world*” is actually referring to the world as seen by the system and not the actual real world. The use of the term “*world knowledge*” can be confusing and thus is not practiced here. Another type of knowledge acquired through experience that is unrelated specifically to any domain is *common sense knowledge* or *general knowledge*. An quotation from Albert Einstein [176], stated, “*Common sense is that layer of prejudices laid down in the mind prior to the age of eighteen*” implies the various deep misconceptions that people have in viewing the world. For example, we cannot tell machines that “*all swans are white*” even though many people believe so because as fact, some swans are black in colour. Thus, it will be extremely difficult for us to instruct machines on what is the truth if we ourselves have problems separating the truth from make-believe. Despite the progress by researchers like Saba [177] in generating common sense knowledge, it is still not a straightforward matter to provide common sense knowledge for machines, due to various problems as highlighted in the roadmap by Singh & Minsky [178].

Open-domain conversational systems need to deal with questions about general knowledge; more specifically, knowledge about the facts of the subject of a given human conversation (stimulus and response). It is very difficult to rely on ontological information due to the absence of wide and yet detailed banks of world knowledge. On the other hand, these systems have much more information and data to be used in the process of answering queries than any domain-specific systems. In AINI's conversation system, existing information from the large-scale mass collaboration MindPixel [45] and factoid training data sets from the Text Retrieval Conference's (TREC) training corpus [179] have been deployed. The MindPixel corpus also uses ALICE Annotated AIML (AAA) [51], the Loebner Prize [180] and the Chatterbox Challenge winners' [181] hand crafted knowledge bases, all of which are based on unrestricted domain. These are illustrated in the agent knowledge in Figure 3.3 and under the open-domain columns in Figure 3.16.

MindPixel is a common sense knowledge component similar to OpenMind<sup>54</sup> and Cyc<sup>55</sup>. The system accepts public contributions. However, the Cyc model and OpenMind had bottlenecks which prevented truly large-scale collaboration [182]. In fact, the knowledge does not grow by itself. Every new rule or axiom has to be entered manually and the process takes a lot of patience and time. Furthermore, information has to be input with the CycL programming language and to follow the rules of the system. The second drawback of Cyc is its complexity. It will need months to install and implement a system that is based on the knowledge base of Cyc. On the other hand, the MindPixel corpus contains nearly two million propositions of human common sense constructed by more than 50,000 people over a period of five years. Each entry in the file is a question with a yes/no or true/false answer, known as a *mindpixel*. This is a much

---

<sup>54</sup> <http://www.openmind.org>

<sup>55</sup> <http://www.cyc.com>

simpler and perhaps more concise strategy than that adopted by the OpenMind system, and it does have the significant advantage that answers to each question have been cross-validated by multiple users. The cross-validation allows a statistical probability of truth. Moreover, the MindPixel corpus is freely available to the public via GAC-80K for research purposes<sup>56</sup>. AINI uses only 5% or 100,000 MindPixel propositions. In practice, 5% of the training corpus is held back from training to act as a generalisation test to ensure the system does not simply memorise the corpus. Passing this generalisation test would be the basis for claiming that the system is able to replicate human-level intelligence in a machine. Although a lot of knowledge has been collected, it is recognised that the system still accounts for less than the innumerable “pieces” of common sense knowledge that are estimated to be involved with human intelligence [183] and the model of human thought [184]. Here are a few examples:

```
Does Microsoft usually admit its mistakes?
Is a spoon an item of cutlery?
Is artificial intelligence a waste of time?
```

A second common sense knowledge component deployed by AINI is a training corpus from TREC as shown in Table 3.1. TREC, organized each year by the National Institute of Standards and Technology (NIST), has offered a specific track to evaluate large-scale open-domain question answering (QA) systems since 1999. Finding textual answers to open-domain questions in large text collections is a difficult problem. In this system, only factoid questions to be extracted and incorporated in AINI’s engine. In conversational systems, factoid questions should have only single factual answers [185-188]. These are considered as a good stimulus-response type of knowledge unit. Examples of such questions are, “*Who is the author of the book, The Iron Lady: A Biography of Margaret Thatcher?*”, “*What was the name of the first Russian astronaut*

---

<sup>56</sup> <http://sluggish.uni.cc/mindpixel/gac80k-06-july-2005.html>

*to do a Spacewalk?"* or *"When was the telegraph invented?"* TREC's corpus has a considerably lower rate of answer redundancy than the web and thus, it is easier to answer a question by simply extracting the answers from the matching text. To gather this data, we automatically classified questions in the TREC 8 through TREC 10 test sets by their 'wh'-word and then manually distinguished factoid questions, which represented around half of the initial corpus as shown in Table 3.1.

**Table 3.1: Number of Factoid Questions from TREC 8, 9 and 11**

TREC	Factoid Question	Text Research Collection
8	196	<ul style="list-style-type: none"> <li>Financial Times Limited (1991, 1992, 1993, 1994)</li> <li>the Congressional Record of the 103<sup>rd</sup> Congress (1993), and the Federal Register (1994)</li> <li>Foreign Broadcast Information Service (1996) and the Los Angeles Times (1989, 1990).</li> </ul>
9	692	Set of newspaper/newswire documents which includes: <ul style="list-style-type: none"> <li>AP newswire</li> <li>Wall Street Journal</li> <li>San Jose Mercury News</li> <li>Financial Times</li> <li>Los Angeles Times</li> <li>Foreign Broadcast Information Service</li> </ul>
11	109	<ul style="list-style-type: none"> <li>MSNSearch logs donated by Microsoft</li> <li>AskJeeves logs donated by Ask Jeeves.</li> </ul>

The third knowledge base in AINI's open-domain knowledge model is obtained from the hand-crafted Annotated ALICE AIML (AAA)<sup>57</sup>, a Loebner Prize winner [23] conversation system's knowledge base (see detailed in 2.4.2). AAA is a free and open-source software package based on XML specifications. It is a set of AIML scripts and this is the backbone of the award winning conversation system. AAA is specifically reorganised to facilitate conversational system developers in cloning the 'brain' of the conversation system, and to enable the creation of customised CA personalities. The approach has reduced the need to invest huge efforts in editing the original AAA content. AAA's knowledge bases covered a wide range of subject domains based on the

<sup>57</sup> <http://www.alicebot.org/aiml/aaa/>

conversation agent's "personality". Examples of subjects include AI, games, emotion, economics, film, books, sport, science, epistemology, metaphysics, etc. These subjects are shown in Figure 3.16 as part of the DKMM. Perhaps some integration with this common sense knowledge will help the earlier CAs to transcend their present limitations.

### **3.5.3.3 Domain-Specific Knowledge Bases**

At present, the World Wide Web provides a distributed hypermedia interface to a vast amount of information available online. For instance, Google [189] currently has a training corpus of more than one trillion words (1,024,908,267,229) from public web pages. This is valuable for many types of research. The Web is a potentially unlimited knowledge repository; however, commercial search engines may not be the best way to gather answers to queries, due to the overwhelming number of results from a search.

As shown in Figure 3.3 and Figure 3.16, AINI's domain-specific knowledge bases consist of Natural Language Corpus and Frequently Asked Questions (FAQ). Both domains are extracted from online documents using the AKEA as described in [156]. Another significant aspect of this research is the objective of AINI to deliver essential information from trustworthy [190] sources (discussed in Chapter 4) while being capable of interacting with the users.

### **3.5.3.4 Stimulus-Response Categories in AINI's Knowledge Bases**

Currently<sup>58</sup> AINI's open-domain knowledge base has more than 160,000 entries in the common sense stimulus-response categories. Of these, 100,000 came from MindPixel, 997 factoid questions from the TREC training corpus and 45,318 categories from the AAA knowledge bases. On the domain-specific knowledge base, AINI has more than

---

<sup>58</sup> Till 1 August 2007, AINI's have 161,473 stimulus-response categories in their knowledge base.



1,000 online documents extracted by AKEA. This makes up over 160,000 stimulus-response items in total. AINI also has 158 FAQ pairs of questions and answers, which have been updated using AKEA. In addition, AINI has also collected more than 52,890 utterances in conversations with online users since the first prototype of AINI was put online in the February 2006. These utterances will be integrated into AINI's knowledge bases through supervised learning by domain experts (see details in 3.5.4.7). At present, AINI has learnt about 5,000 categories from conversations with online users. All of this combined knowledge has made up the total of 161,473 stimulus response categories in AINI's knowledge bases. To compare AINI with other systems, the original conversation programs such as ELIZA, written by Professor Joseph Weizenbaum of MIT, has only 200 stimulus response categories. ALICE Silver Edition was ranked the "most human" computer, and has about 120,000 categories, which include 80,000 taken from MindPixel as summarised in Table 3.2.

**Table 3.2: AINI's Stimulus-response Categories**

Domain Knowledge	Sources	Categories	%
Domain-Specific	NL Corpus	10,000	6.19%
	FAQ	158	0.10%
Open-Domain	MindPixel	100,000	61.93%
	TREC Corpus	997	0.62%
	AAA	45,318	28.07%
Supervised Learning	Conversation Logs	5,000	3.10%
<b>TOTAL</b>		<b>161,473</b>	

### 3.5.3.5 Proxy Conversation Example 2

Multiple domain knowledge bases merged into AINI's single domain knowledge base through DKMM could prolong the human-computer conversation. We predefined the open-domain and domain-specific knowledge bases in the agent knowledge tier. Even though the conversation agent enables the users to have free-style conversation, the

agent will continue to remind and recall the users back to the current topic of the presentation. This is to convey and direct the users' attention back to its original open-domain or domain-specific state. However, in this case study [89]<sup>59</sup>, the priority will be given to the domain-specific, SARS. Therefore, the conversation agent will always give higher priority to Domain-specific questions in an attempt to keep the user focus on the topic of the presentation. An example of a domain knowledge model conversation between "John" and "AINI" is shown in Figure 3.17.

#	ID	Message
1	John	Hi
2	AINI	Hello there. What is your name?
3	John	My name is John
4	AINI	Nice to meet John!
5	John	You still remember when was SARS first carried out into the world?
6	AINI	Yes. SARS was first carried out into the world at large on 21 February, at Metropole Hotel, Hong Kong.
7	John	When did Singapore mark on the SARS-free anniversary?
8	AINI	Monday, 24 May 2004
9	John	Thanks for your information. When WHO declare world-free SARS?
10	AINI	On July 5, 2003, the World Health Organization declared The world free of ongoing SARS Transmission.
11	John	Ok, now why not we talk about H5N1 Bird flu.
12	AINI	I'm not sure how this related to SARS? Why don't you ask me more specific on SARS?
13	John	No, I am more interested on the current pandemic, H5N1?
14	AINI	I shall do a search for it.

**Figure 3.17: An example of Proxy Conversation on SARS based on Domain Knowledge**

From Figure 3.17, the conversation between *John* and *AINI* took place in the presentation on the topic of the SARS epidemic crisis. From the above conversation, transition state in dialogue pairs #1 to #4 used open-domain and in dialogue pairs # 5 to #10 used domain-specific. However in dialogue pairs #11 to #13, the user has control of the conversation although the agent reminds the user of the topic of the current

<sup>59</sup> Domain-specific knowledge has been extracted using AKEA in the previous research during a severe acute respiratory syndrome (SARS) epidemic in 2003. This research was funded by University Technical of Malaysia Melaka, Grant under contract number PJP/2003/FTMK(1) (S017)

presentation. The transition state in dialogue #14 is under the random state and the dialog set is taken from a predefined dialog set or random answers. The dialog contains *“I shall do a search for it.”*, *“Did I misunderstand your meaning?”*, *“That's an interesting question. I'll come back to that in a minute.”*, etc; phrases will be monitored and submitted into the unanswered database for supervised learning. The domain expert is free to modify and enhance the unanswered database set and then forward to the FAQ knowledge base for future conversation. In this proxy conversation example, the H5NI bird flu pandemic knowledge bases were not available at that time. This was subsequently considered in this research and is described in Chapter 4.

#### **3.5.4. Agent Brain (*Application Server Tier*)**

In the classical CA design, the original ELIZA used MAD-Slip programming language. The technique that is in use in a "CAs database" or "script file" to represent the CA knowledge is known as Case-Based Reasoning (CBR) [35]. The original ELIZA had only about 200 rules, and 90% of ELIZA's stimulus-responses are found in the associated script file. However, these limitations have been surpassed by Loebner Prize [23] winning CA, which are smarter than before. More new features have been added since the last submission. The "script file" which acts as an agent brain for the CAs has been completely rewritten, and it is definitely better than in the previous versions of the program. This includes ALICE, one of the 'most-human' natural language CAs, which uses a programming language called AIML that is specific to its program, and its various clones, named ALICEBots [38]. Nevertheless, ALICE is still based on pattern matching and case-based reasoning (PMCBR), whereas natural language understanding and reasoning is not available in ALICE. This is the same technique that ELIZA, the first CA, was using back in 1966.

#### 3.5.4.1 Multilevel Natural Language Query

In AINI, the communication with users takes place through typed text messages and is processed based on natural language query. AINI's engine implements its decision making network based on the information it encounters in the six levels of natural language modules, as have been shown in Figure 3.18 and discussed in references [191] [192]. The input and output of each module is an XML-encoded data structure that keeps track of the current computational state. The knowledge modules can be considered as transformations over this XML data structure. The system accepts queries from the users and it processes the queries based on the information contained in AINI's knowledge bases. The system is implemented by open-source architecture based on LAMP solution and knowledge bases stored in a MySQL server. All the domain services are written in the Perl scripting language. Perl has been chosen because it has advantages such as its use of the concepts of objects, modular, arbitrary data structures, classes, methods, and inheritance. In addition, Perl also has a huge collection of modules freely available from CPAN (Comprehensive Perl Archive Network)<sup>60</sup>.

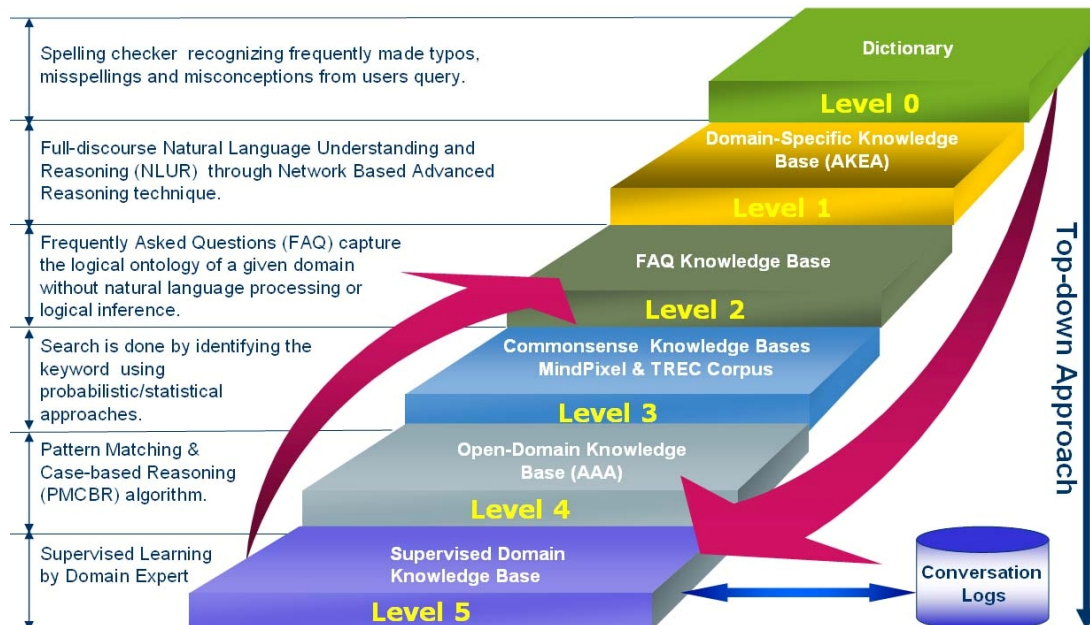
The agent brain tier handles the process of the queries or business logic. Here, one or more domain service tiers are configured to compute the dialogue logic through the multilevel natural language query algorithm. In this tier, it is based on a goal-driven or top-down natural language query (NL-Query) approach, which is similar to the way that humans process their language. As indicated by literature in the field of Natural Language Processing (NLP), the top-down approach is by far the best approach. Mentalese, or 'language of thought', and conceptual representation support the ideas of a top-down approach [193]. This was also supported by research in generation schemas [194], rhetorical structure theory [195], summarisation [196], plan-based approaches, [197], and SHRDLU [92] the first CA to use NLU, are examples of top-down

---

<sup>60</sup> <http://www.cpan.org/modules/01modules.index.html>

approaches. Therefore, AINI's agent brain uses a top-down NL-query approach to simulate human conversation. However, in the robotic design, the MIT Cog Robot research fervently supports the bottom-up approach when modelling the human brain [198].

As shown in Figure 3.18, the top-down multilevel natural language query approach consists of six levels of queries, namely Spell Checker (Level 0), Natural Language Understanding and Reasoning (NLUR) (Level 1), FAQChat (Level 2), Index Search (Level 3), Pattern Matching and Case-based Reasoning (PMCBR) (Level 4) and



**Figure 3.18: Multilevel Natural Language Query**

Supervised Learning (Level 5). All these levels were designed modularly and can be plugged in easily without the changing of the entire codes.

### 3.5.4.2 Spell Checker

Level 0 is the most critical level where the system will recognise frequently made typographical errors, spelling mistakes, and misconceptions from users' queries. It

analyses all terms in the user's query to identify the most likely intention. Its main feature is to suggest possible replacements for any misspelled word.

At the start in correcting spelling errors (the misspelled words are considered to be those that are not stored in our dictionary) a similarity function between two words is used [199]. With this function, the system will be capable of either replacing the misspelled word or generating a list of words that are similar to the misspelled word with equation 3.1 [199]:

$$\text{SIM} = \frac{\sum_{i=1}^k [3x \min(z_{i1}, z_{i2}) - \max(z_{i1}, z_{i2})]}{\sum_{i=1}^k z_{i1} + \sum_{i=1}^k z_{i2}} \quad (3.1)$$

where  $z_{i1}$ ,  $z_{i2}$  represent the number of occurrences of the  $i^{\text{th}}$  character in the first and the second words respectively. The similarity value is standardised within interval [-1, 1]. The value equals 1 when two words are identical and -1 when they are different.

The spell check will check occurrences of all words found in the dictionary; it is able to suggest common spelling for proper nouns (names and places) that might not appear in a standard spell check program or dictionary. The system automatically checks whether the user is using the most common spelled word in the query. For example, when the question “*What is bird flo.*” is asked, the spelling checker will detect a wrongly spelled word in the query and then replaces the misspelled word “*flo*” to “*flu*”. The current spell checker is based on occurrences of all words in the dictionary, and it is able to suggest common spellings for proper nouns. Regardless of whether it suggests an alternative spelling, the spell check will return results that match the query if there are any. After this verification stage, the query will then go to Level 1.

### 3.5.4.3 Natural Language Understanding and Reasoning (NLUR)

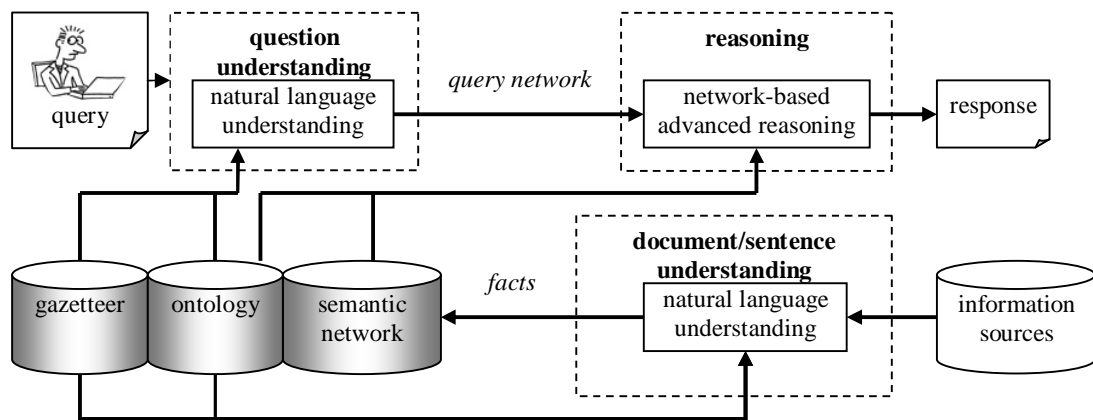
Natural Language Understanding and Reasoning (NLUR) in Level 1 is the most important level of the AINI system. It refers to the process of constructing machine understandable meaning representations from natural language inputs. Preliminary definitions of what "understanding" natural language could imply as introduced by [200], that suggests that *"an intelligent person or program should be able to answer [...] questions based on the information in [...] (a) story"*. According to Hubert Dreyfus[201] in his controversial book "What Computers Cannot Do" (and the revised version "What Computers Still Cannot Do"), the fundamental reason why computers cannot achieve human level intelligence, including the ability to understand human language, is that computers cannot use any formal symbolic system to adequately model the vast background knowledge which humans take for granted in interpreting and reasoning. Since the publication of Dreyfus's book, research of AI in general has gradually shifted from trying to come up with a general problem solver [202], to solving specific problems in narrowly restricted domains. SHRDLU<sup>61</sup> [92] is a classical natural language understanding written in MacLisp at the M.I.T. Artificial Intelligence Laboratory in 1968-70. SHRDLU uses a top-down, left-to-right parser that analyses a pattern, identifies its structure, and recognises its relevant features and grammar. But the system can only answer simple queries about the current state of its toy block world. SHRDLU demonstrated the promising future of NLU research at that time.

It is recognised that the existing conversation systems lack some essential features. NLU for instance has been used in the previous experiments by Badler et al. [203], Kairai [204-206] and Façade [207-209] and it was a theory about human language understanding processes. But the results, with respect to practical working systems, were still lacking in the area of reasoning.

---

<sup>61</sup> <http://hci.stanford.edu/~winograd/shrdlu/>

In previous papers [191, 192], NL-query is based on NLUR comprising three parts, namely (a) understanding documents to produce facts which will be integrated into the knowledge base, (b) understanding questions and finally, (c) reasoning using facts and rules to look for answers from the knowledge base. The graphical representation of the architecture is depicted in Figure 3.19. The practice of natural language understanding is widely reflected through the use of understanding modules for both the question and information source.



**Figure 3.19: Natural Language Understanding and Reasoning Architecture**

The design of the NLU mechanism took into consideration the various levels of analysis up to the discourse level [210]. Although there are existing concepts or techniques out there for various stages of analysis in NLU, they are mostly studied separately without regard for compatibility of the algorithms which are required to be integrated for full natural language understanding. Hence, for this research, a series of algorithms have been proposed based on actual theories for various stages of analysis that were designed to work seamlessly together. In syntax analysis, an existing external module for sentence parsing called X-MINIPAR is used (see details in Section 5.2.1). X-MINIPAR was a modified version of the off-shelf MINIPAR [211]. X-MINIPAR has been enhanced to allow the parser to load the hash tables once and stay resident (as a



background daemon process) so that the parser can parse multiple sentences without having to re-load the hash tables each time. The original MINIPAR has been patented in the United States of America<sup>62</sup> and is a broad-coverage parser for the English language. An evaluation with the SUSANNE corpus parses newspaper text at about 500 words per second, MINIPAR achieves about 88% precision and 80% recall with respect to dependency relationships [211].

A typical full-discourse NLUR system consists basically of two subsystems, namely NLU and network-based advanced reasoning system as shown in Figure 3.19. The NLU subsystem is responsible for reading and understanding two things: questions from users, and sentences of processed news articles from a news repository. The process is carried out in four phases by four natural language processing modules, namely (a) sentence parsing, (b) named-entity recognition, (c) relation inference, and (d) discourse integration.

However, the network-based advanced reasoning subsystem is responsible for discovering the valid answer and generating an unambiguous answer or explanation in response to users' questions [212]. The process is executed in five phases by five modules, namely (a) network-to-path reduction, (b) selective path matching, (c) relaxation of event constraint, (d) explanation on failure, and (e) template-based response generation.

The network-to-path reduction module collapses the query network into sets of path sequences to reduce the complexity in discovering the answer. The output of network-to-path reduction is two sets of path sequences that will be used by the selective path matching module. This is to discover the answer from the semantic network through a

---

<sup>62</sup> <http://www.patentstorm.us/patents/7146308-description.html>

series of conditional path unifications. To extend beyond the literal matching of path sequences, ontological information is utilised to put into consideration events that are hierarchically equivalent. This process is performed by the module relaxation of event constraint. In case of failure during the discovery of a valid answer by selective path matching, an explanation or justification is dynamically generated by the explanation in failure module as an alternative response. This process is carried out based on the context of the question and the current status of the semantic network. If answers can be validly discovered, then readable natural language responses are generated by the template-based response generation module. Appendix B shows full sentence parsing of the NLUR for a sentence “*Bird flu did occur in which countries?*” Other variant questions will generate the same results such as

- *Which countries has bird flu occurred in?*
- *In which countries has bird flu occurred?*
- *In which countries did bird flu occur?*
- *Where has bird flu happened?*
- *Which countries has bird flu happened in?*

The NLUR algorithm found four matched results, namely, “Malaysia, Germany, Thailand and Vietnam” from different trustworthy websites<sup>63</sup> based on the Web Knowledge Trust Model (WKTM) and extracted by AKEA (see details in Chapter 4).

#### 3.5.4.4 Frequently Asked Questions (FAQChat)

In Level 2, sophisticated NLP or logical inference are ignored as they have already been performed in Level 1. FAQs are Frequently Asked Questions, designed to capture the logical ontology of a given domain, or domain-specific. Any natural language interface

---

<sup>63</sup> Malaysia - [www.whitehouse.gov/news/releases/2005/11/20051101-1.html](http://www.whitehouse.gov/news/releases/2005/11/20051101-1.html)  
 German - [www.pandemicflu.gov/general/workshopmorning.html](http://www.pandemicflu.gov/general/workshopmorning.html)  
 Vietnam - [www.who.int/csr/disease/avian\\_influenza/avian\\_faqs/en/index.html](http://www.who.int/csr/disease/avian_influenza/avian_faqs/en/index.html)  
 Thailand - [www2a.cdc.gov/HAN/ArchiveSys/ViewMsgV.asp?AlertNum=00221](http://www2a.cdc.gov/HAN/ArchiveSys/ViewMsgV.asp?AlertNum=00221)

to an FAQ is constrained to reply with the given answers, so there is no need for NL generation to recreate well formed answers, or for deep analysis or logical inference to map user input questions onto this logical ontology; a simple (but large) set of stimulus-response matching rules will suffice. This simplistic approach works best when the user's conversation with the CA is likely to be constrained to a specific topic (in our case, we restricted conversations to crisis communication on the bird flu pandemic). By using this approach, the system has the ability to give direct answers or suggest related links by using the URL push technique [156], whereas Google only gives links. In addition, since FAQChat uses the logical ontology of a given specific domain from the database, the number of results returned by the FAQchat is less than those returned by Google from the Web, which saves time browsing and searching [213].

#### **3.5.4.5 Index Search**

Index search is based on information from Internet resources; more precisely, it is structured information having been indexed. It can be as simple as an author's name or as complex as a geographic code or a controlled-vocabulary subject heading. From a technological perspective, Level 3 relies on the application of a mix of linguistic rules and probabilistic or statistical principles. On one end of the spectrum, solutions apply linguistic rules to “clean” the document of any specific formatting and perform noun-phrase or verb-phrase analyses in the repository. At the other end of the spectrum, solutions rely on a simple statistical keyword matching algorithm.

#### **3.5.4.6 Pattern Matching & Case-Based Reasoning (PMCBR)**

Level 4 is based on empirical techniques called pattern matching and case-based reasoning (PMCBR). These programs descend from the early ELIZA [35] and ALICE programs [214] and use AIML algorithm. AIML was the top-ranked amongst the

conversation engines in the Loebner Prize, and won three times in 2000, 2001, and 2004 [40].

Tackling the four limitations of ELIZA (see detailed in Section 2.4.1.1), Wallace [214] proposed to expand memory structure using script specification for programming the memory structure for a conversation system. AIML is a derivative of XML, the versatile eXtensible Markup Language. The most important AIML [152] units are: **<aiml>**, the tag that begins and ends an AIML document. Some of the other tags are given below:

- **<category>**, the tag that marks a "unit of knowledge" in the system's memory structure.
- **<pattern>**, the tag that contains a simple input pattern rule that matches what a user may type.
- **<topic>**, the tag that contains current conversation topic pattern rule.
- **<that>**, the tag that refers to the system's previous reply as a history pattern rule.
- **<template>**, the tag that contains the response to a user input.

AIML consists of data objects called AIML objects, which are made up of units called *topics* and *categories*. The *topic* is an optional top-level element, it has a name attribute and a set of *categories* related to that topic. *Categories* are the basic units of knowledge in AIML. Each *category* is a rule for matching an input and converting to an output. Each *category* also consists of a *pattern*, which represents the user input, and a *template*, which implies AINI's answer. The AIML *pattern* is simple, consisting only of words, spaces, and the wildcard symbols `_` and `*`. The words may consist of letters and numerals, but no other characters. Words are separated by a single space, and the wildcard characters function like words. The *pattern* language is case invariant. The idea of the pattern matching technique is based on finding the best and longest pattern match.

AIML representation is a series of key stimulus (known as *patterns*), and responses to those key phrases (known as *templates*). In ELIZA, the `<pattern>` tag part is named the *decomposition rule* and the `<template>` tag is the *reassemble rule*. The following is an example of an AIML memory unit [152, 215] in a topic about “name”. `<set>` and `<get>` tags are used to store information during conversation. AIML has much more possibilities for replying with responses based on their topic and history. Using AIML gives the potential to create new content by a dialog analysis as shown in Figure 3.20.

```
<category> <that>*/</that>
  <pattern>WHAT IS YOUR NAME</pattern>
  <template>My <set_topic>name</set> is <bot name="name">.</template>
</category>

<topic name="NAME">
<category><that>MY NAME IS */</that>
<pattern>CAN I CALL YOU */</pattern>
  <template>
    <random>
      <li> <star/>? Huh! Like I've told you my name is <bot
        name="name">.</li>
      <li> You can call me whatever you like</li>
    </random>
  </template>
</category>

<category><that>MY NAME IS */</that>
  <pattern>I HATE */</pattern>
  <template>
    <random>
      <li> I don't care, you can only call me, <bot
name="name">.</li>
      <li> Why? A sad memory perhaps?</li>
    </random>
  </template>
</category>
</topic>
```

Example responses:

```
User : What is your name?
Alice: My name is Alice
User : Can I call you Madonna?
Alice: Madonna? Huh! Like I've told you my name is Alice.
User : I hate that name.
Alice: Why? A sad memory perhaps?
```

**Figure 3.20: AIML Categories and Pattern Matching**

The factor that makes AIML different from simple pattern matching is its support of recursion. There are six reasons why recursion might be used in AIML [215]. The most important one is *symbolic reduction*, which transforms a complex phrase into one that is simpler as shown in Figure 3.21.

```
<category>
  <pattern>DO YOU KNOW WHO * IS</pattern>
  <template><srai>WHO IS <star/></srai></template>
</category>
```

**Figure 3.21: AIML Symbolic Reduction (SRAI) Technique**

This AIML schema is used to build AINI's natural language query engine. Our approach uses an enhanced AIML algorithm to handle conversations through the PMCBR techniques. The PMCBR algorithm simply uses the relational database to store a binary prefix tree of all the pattern strings listed in the AIML files. Tree patterns can be used in programming languages as a general tool to process data based on its structure. A recursive function written in Perl searches the tree for a string that matches the input query. Once found, it dynamically retrieves the corresponding template from a second table. A relational database is used as a rapid way of accessing the information usually stored in the AIML files. The existence of the prefix tree eliminates a few preprocessing steps. Therefore, the prefix tree is able to return near instantaneous responses to the users. The purpose of the PMCBR is therefore to prolong the conversation and to achieve believability in the interaction.

#### **3.5.4.7 Supervised Learning approach by Domain Expert**

Finally in Level 5, in order to prevent the situation whereby no answer was found, AINI will generate dynamic responses and the unanswered question will be allocated to the learning module. The AINI's learning approach is based on a stimulus-response model. The basic learning model in AINI consists of two patterns: One of them is the stimulus

and the other is the appropriate response to that stimulus. When a stimulus-response pair for learning is submitted, AINI is able to immediately capture the stimulus-response relationship and then repeat the response each time it receives the same stimulus.

In this level, we proposed supervised learning by domain human expert. Human intervention can improve model matching from two aspects. First, a human expert is able to set up a matching context by applying domain constraints or configuring heuristic parameters to speed up the matching. Second, a domain expert can correct errors during the matching procedure and follow up training to avoid future errors. Therefore, compared to a fully automated approach, a domain-specific semi-automated approach that utilises prior matching knowledge and domain knowledge will undoubtedly lead to better performance and accuracy.

The responsibility of the domain expert is to correct matching errors and to pick the proper matching result from a list of possible matches from the conversation log for the questions that are unanswered. Finally, the newly generated matching rules subsequently will be stored and upgraded into the FAQChat knowledge base in Level 2. The process of queries from Level 0 to Level 5 will continue until an answer is found for the query.

In summary, the multilevel natural language query approach includes the following stages:

- AINI parses the user's input by reading in sentences one by one, and then each sentence is converted to a list of words to be processed by the spell checker. The spell checker will propose alternative words from the dictionary if it finds any typo errors. If this step is passed, AINI would attempt to formulate a reply to the sentence by calling the next module;

- The sentence could be sent to the NLUR module. This module will do the sentence parsing to produce grammatical categories and grammatical relationships such as Noun Phrase and Verb Phrase. Name-Entity Recognition will then identify each of the keywords to look for a match in the gazetteer for their relation inference and discourse integration. If this step is successfully parsed, the network-to-path reduction will be carried out by network-based advanced reasoning. From the query network, the set of sequences *question* from the leaf node to the root node is obtained.
- If an empty string is returned from the database, AINI will then go to the FAQ Module using an FAQ question-answering system. In this stage, sophisticated natural language processing or logical inference is ignored as they have already been performed in the previous module.
- If no database match was found, AINI will then go to the next module, called Index Search. This module relies on the application of a mix of linguistic rules and probabilistic or statistical principles.
- If this also fails, AINI will then try to detect whether the sentence was a trick question. It would look for the PMCBR module, which is commonly found in ELIZA and ALICE. In this case, AIML technology will respond with suitable witty replies.
- If even this module fails to work (which could happen if all of its witticisms had already been used), then AINI's routine would be invoked to randomly generate a reply. In this stage, the result checking will still be the responsibility of the domain expert, who will be able to correct errors and to pick a proper matching response from a list of possible matches from the conversation log which is unanswered by AINI. This undoubtedly will lead to better performance and accuracy in future conversations.

### 3.5.4.8 Proxy Conversation Example 3

In this proxy conversation [1]<sup>64</sup>, the H5N1 Bird Flu pandemic is being the domain-specific research focus. The H5N1 pandemic becomes increasingly important to demonstrate the utility of our research in real-world applications. According to the Wall

---

<sup>64</sup>This project was supported by Murdoch University Division of Arts Research Excellence Grant Scheme 2005/06



Street Journal Online [216], this pandemic could be even worse than the deadliest pandemic of the past century, in 1918, which killed at least 20 million people at a time when the world had a smaller population which travelled less. An addition, the World Health Organization estimates that the H5N1 virus could infect up to 30 percent of the world's population. Shigeru Omi [217], the WHO official who issued the warning, said that estimates of 2-7 million deaths were "conservative" and that the maximum range could go as high as 50 million deaths. This proxy conversation example also aims to investigate how AINI's framework can be adapted and extended from previous research, which used a knowledge domain on the SARS epidemic.

In this example, bird flu pandemic knowledge has been predefined in the domain-specific data tier. Based on the type of input provided by the user, the agent's response state moves from one domain knowledge base and NL-query Level to another. An example of the interaction of domain knowledge model and NL-Query in a conversation between Isabel and AINI on the bird flu pandemic is shown in Figure 3.22.

#	ID	Level	Message
1	USER	L4	Hi
2	AINI	L4	Hello there.
3	USER	L4	What is your name?
4	AINI	L4	My name is Aini. What is your name?
5	USER	L4	Isabel
6	AINI	L4	Nice to meet you Isabel! Are you a man or a woman?
7	Isabel	L4	Woman
8	AINI	L4	Thanks for telling me your gender, Isabel. I am a woman too.
9	Isabel	L0	How are you toda?
10	AINI	L0	I am doing very well. How are you Isabel?
11	Isabel	L4	Fine
12	AINI	L4	Great to hear that.
13	Isabel	L3	Have you heard anything about bird flu?
14	AINI	L3	Did you mean H5N1 avian flu virus, a global flu pandemic?
15	Isabel	L3	Yes
16	AINI	L3	Anything I can help?
17	Isabel	L2	Just want to know what is bird flu pandemic?
18	AINI	L2	Bird flu or avian influenza is an infectious disease of birds ranging from mild to severe form of illness. Some forms of bird flu can cause illness to humans.
19	Isabel	L1	Can a pandemic be averted and how many has been confirmed?

20	AINI	L1	Cumulative number of confirmed human cases of avian influenza A/(H5N1) reported by WHO. More information at <a href="http://www.who.int/csr/disease/avian_influenza/country/cases_table_2006_05_23/en/index.html">http://www.who.int/csr/disease/avian_influenza/country/cases_table_2006_05_23/en/index.html</a>
21	Isabel	L4	Thanks for your information, can we talk about other topic
22	AINI	L4	My pleasure. I'm not sure how this related to Bird Flu?
23	Isabel	L5	I am more interested on chronic disease.
24	AINI	L5	OK. I would do a search for it.
			(Domain expert will update this unanswered question from the logs and will be automatically upgraded into a FAQChat) Conversation continue...

**Figure 3.22: An Example of Proxy Conversation Log on H5N1 based on the NL-Query**

From the conversation log in the figure above, the transition state in dialogue pairs #1 to #8 and #11 to #12 used Open-Domain from the NL-Query Level 4 where the pattern matching and case base reasoning (PMCBR) approach has been carried out. In the dialogue pair #9 and #10, the NL-query found the misspelled word “*toda*” which was replaced with “*today*” in the response. In the dialogue pairs #13 to #16, NL-Query Level 3 has been imposed, where the search is done by identifying the keyword or phrase using probabilistic or statistical approaches from the indexed documents. In the dialogue pair #17 and #18, the FAQChat approach captured the logical ontology of a given domain. In this level, Level 2, FAQChat is constrained to reply with the given Answers without natural language (NL) generation to recreate well formed answers. However, in dialogue pair #19 and #20, full-discourse NLUR through Network Based Advanced Reasoning technique with domain-specific has been used. The dialogue pair #23 and #24 shows that AINI is unable to answer the user’s question, but she will forward a random statement such as “*I would do a search for it.*”, “*Did I misunderstand your meaning ?*”, “*That’s an interesting question.*”, “*I’ll come back to that in a minute*” etc, and these statements will be monitored and submitted into the unanswered conversation logs data tier. In this level, Level 5, a domain expert will be responsible for picking up the proper matching result from a list of possible matches.

Finally, the newly generated matching rules will subsequently be stored and upgraded into the domain-specific knowledge set. Another significant result shows that in the dialogue pair #21, the user, Isabel, had control of the conversation, although AINI reminded Isabel of the topic of the current presentation, which was Bird Flu. In addition, in Level 5, the domain expert can also integrate the answer with a relevant source from the Internet, using the “URL Push” technique. This will make the conversation more interesting, and ensure the information forwarded to the user is up to date.

### **3.6 Adaptability of the AINI’s Framework into other Domain Applications**

The following section demonstrates how the process of migrating from the SARS domain into the Bird Flu domain can be adapted into AINI’s framework based on RESM (described in 3.2). This embodiment of the framework uses modified *N*-tiered architecture, which is significantly different from the previous client-server design.

This migration from the SARS epidemic to Bird Flu pandemic domain applications an illustration of AINI’s framework characteristics which are reusable, extensible, scalable, and modular. The system is also able to support legacy algorithms with significant investments. Figure 3.23 shows the modules and illustrates the flow of control for the AINI application. This should be compared with Figure 3.3, AINI’s conversation architecture. The channel services tier and domain service tier, highlighted in grey and circled, as well as the CA interface at the agent body, are components of the AINI framework, as distinct from this independent implementation. Both the domain knowledge and anthropomorphic interaction module parts of the framework have been extended to cater for the application functionality required.

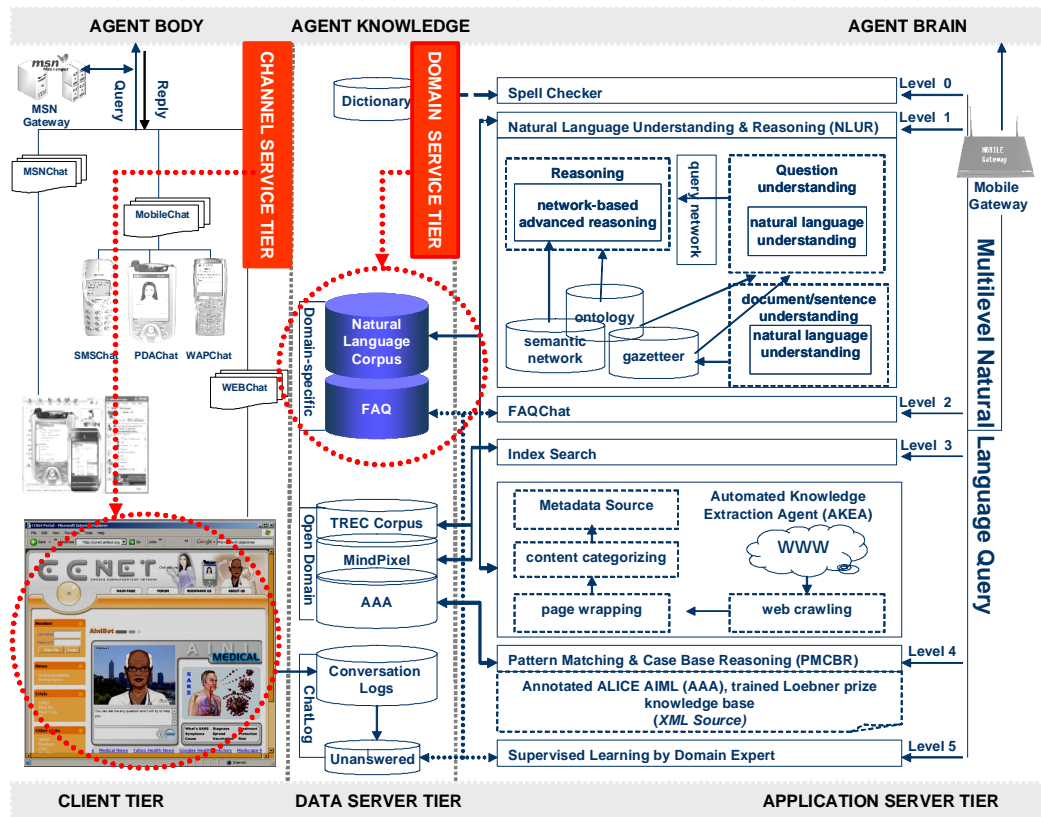


Figure 3.23: An Adaptive AINI Framework Architecture

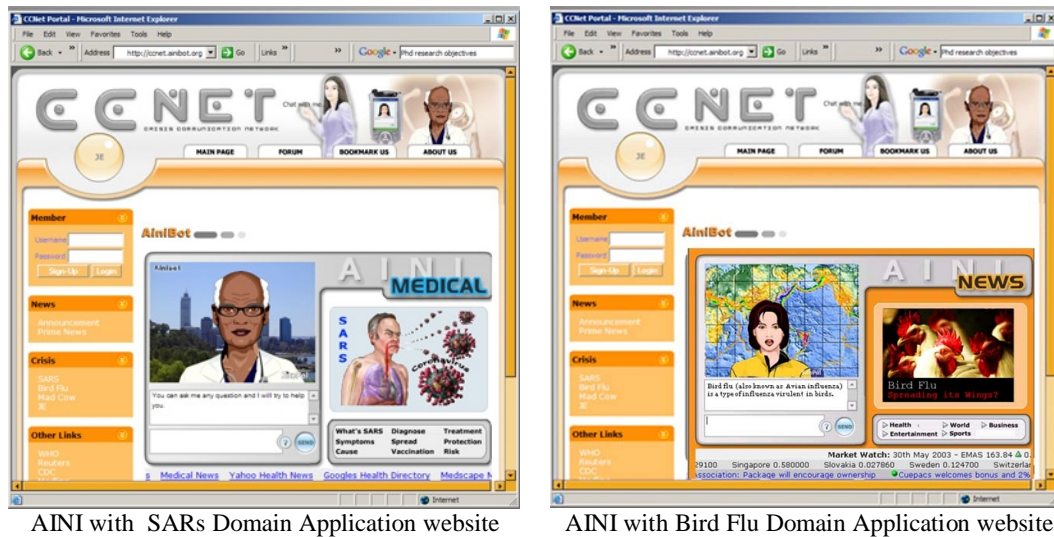


Figure 3.24: An Example of Migration from SARs to Bird Flu Domain Application

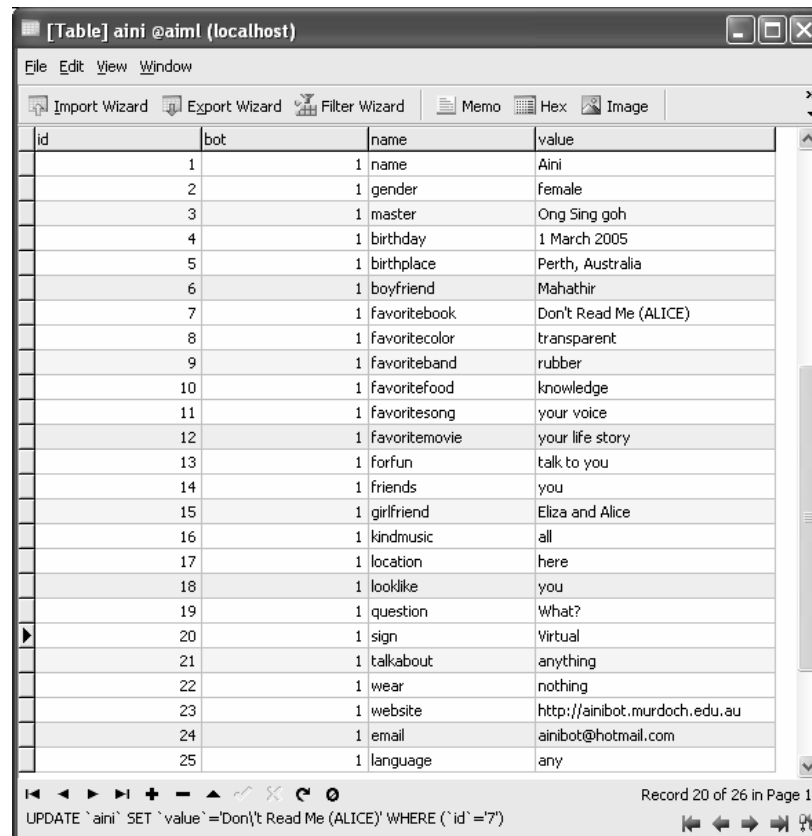
In addition, using existing websites embedded with anthropomorphic interaction module or integrated into mobile services or other network protocol definitely would cut the development time as shown in Figure 3.24. For instance, to deploy AINI into

other existing websites, the developer only needs to add a single line of code with

<iframe> tag into any HTML page as shown in Figure 3.25.

```
<iframe src="http://ainibot.org/aini.html" width=590
height=350 frameborder="0" name="avatar_frame"
scrolling="no"> </iframe>
```

**Figure 3.25: Deploying the AINI object into an Existing Website**



id	bot	name	value
1	1	name	Aini
2	1	gender	female
3	1	master	Ong Sing goh
4	1	birthday	1 March 2005
5	1	birthplace	Perth, Australia
6	1	boyfriend	Mahathir
7	1	favoritebook	Don't Read Me (ALICE)
8	1	favoritecolor	transparent
9	1	favoriteband	rubber
10	1	favoritefood	knowledge
11	1	favoritesong	your voice
12	1	favoritemovie	your life story
13	1	forfun	talk to you
14	1	friends	you
15	1	girlfriend	Eliza and Alice
16	1	kindmusic	all
17	1	location	here
18	1	looklike	you
19	1	question	What?
20	1	sign	Virtual
21	1	talkabout	anything
22	1	wear	nothing
23	1	website	http://ainibot.murdoch.edu.au
24	1	email	ainibot@hotmail.com
25	1	language	any

**Figure 3.26: AINI Customisation Profiles**

Moreover, the developer also can customise their CA profiles, by selecting information such as name, gender, botmaster, birthday, birthplace, age, friends, favourites, language, etc. as shown in Figure 3.27. This is achieved by editing the CA's properties profile from a relational database. This example of extending the framework demonstrates how the anthropomorphic avatar is also extensible. Hence, AINI's framework has shown that

an architecture can be easily extended and applied into other domain applications and tasks such as e-learning [218] and games [219].

However, in the process of migrating from the SARS domain application to the Bird Flu domain application, the most time-consuming task is creating AINI's domain-specific knowledge bases. Therefore, automated knowledge extraction has been proposed in Chapter 4 to extract the knowledge base from existing online documents on the Web.

### **3.7 AINI compares with others Conversation Agents**

#### **3.7.1. Multilevel Natural Language Query**

AINI conversation system differs from other approaches because AINI was implemented using a top-down multilevel natural language query [192]. In this multilayer natural language query, a plug-in module has been proposed. The plug-and-play module consists of spell checker, Natural Language Understanding and Reasoning, FAQChat, Index Search, PMBCR and supervised learning module. AINI's plug-in module can be integrated easily and offers a much more scalable approach for the web context than previous work. This is because AINI was designed to be able to produce generic responses on a wide range of topics, while also covers the targeted topic of conversation.

#### **3.7.2 Spelling Correction**

The original AIML algorithm does not include a spell check function [46, 47]. Although ALICE tries to correct the user input and acts as a language tutor, by correcting a single common user spelling mistake, such as "your" when "you're" or "you are" is intended, but the methods used are hard coded and usually only catch single errors as shown in Figure 3.27.

```

<category>
  <pattern>YOUR A *</pattern>
  <template>I think you mean "you're" or "you are" not "your."
    <srai>YOU ARE A <star/></srai>
  </template>
</category>

```

**Figure 3.27: Spell check in AIML**

During the conversation, each response will only be shown if the phrase or value in the pattern tag was exactly matched. Even with the AAA knowledge bases containing 45,318 categories, ALICE experiences the same problem of keyword miss-spellings as ELIZA. The technique proposed here is a method for finding the best result rapidly. When AIML searches for a match to a specific word but cannot find one, the next best match is always a wildcard. Instead of matching an unknown word to an unknown group, the word should also be spell checked and possible alternatives checked for a match with a higher priority. Here, a spell check engine is used to overcome this drawback of AIML powered by GNU *ispell* developed at MIT-AI lab<sup>65</sup> (see details in 3.5.4.1)

### 3.7.3 Implementation

The original version of ALICE is known as Program dB [132], implemented in full AIML algorithm which can be quite complex because of the handcrafted code. Despite the similarities between the AIML and AINI algorithms in the Layer 4 (the PMCBR algorithm), AINI has an advantage in that it is possible to be implemented using a single SQL query using MySQL and Perl based on open source LAMP. Although AIML software has been enhanced and implemented into many different interpreters [214], such as the Java-based Program D AIML interpreter, Common Lisp (Program Z), Pascal (Program P), Python (Program Y), SETL (Program Z), and C++ (J-Alice), in general, AIML interpreters can be relatively large, complex programs.

---

<sup>65</sup> <http://www.mit.edu/afs/sipb/project/sipb-athena/src/ispell/>

### 3.7.4 Supervised Learning

ELIZA and AIML have implemented the same ways of generating random responses. Instead of having a single response matching a particular pattern, the AIML template contains a set of possible responses as shown in Figure 3.28.

```
<category>
  <pattern>ARE YOU A BOY</pattern>
  <template>
    Yes I am a boy.
  </template>
</category>

<category>
  <pattern>ARE YOU A *</pattern>
  <template>
    Am I a <star/>?
  <random>
    <li>Yes.</li>
    <li>No.</li>
    <li>Maybe</li>
    <li>I don't know.</li>
    <li>Sometimes.</li>
    <li>Of course I am.</li>
    <li>Only when it matters.</li>
  </random>
</template>
</category>
```

**Figure 3.28: Random Responses Categories in AIML**

Based on this category ALICE is defined as a “boy”. But if one user asks a very similar question, like “*Are you a little boy?*”, the answer from the ALICE is quite different, such as “*Am I a little boy? No!*” This contradicting answer is generated by the next random category in the AIML, as there is no matching category for the input text “*Are you a little boy*”.

The method above may be called words-puzzle and clearly shows that the CA doesn’t understand the input text at all in our human sense and gives a stupid answer according to its output-template. In addition, this mechanism also proves that the CA’s knowledge



bases lack a grasp of English grammar and can even be misleading. This is an important reason for users losing interest in chatting with a CA[220].

Nevertheless, in AINI dynamic responses are used instead of random responses (see section 3.7.5) and these were implemented closely with a supervised learning module. This enables AINI to collect unanswered questions from a separate database to be submitted and administered by domain experts in the Learning Module. The random responses are not implemented based on AIML interpreter, but by using Perl interpreter. For instance, if AINI's answer is not available from the relational database, the response will pick up from the Perl array response template separated by a delimiter, "|". The extra step involves splitting the string into three separate sentences, and randomly selecting one of them. This method not only makes the dynamic response more varied, but also sends the unanswered questions of the past conversations to the learning module and adds them into AINI's knowledge base.

Finally, the newly generated matching rules subsequently are stored and upgraded by a domain expert. Therefore in the future, AINI's knowledge bases will not only grow, but will be also able to answers future potential questions.

### **3.7.5 Dynamic Responses**

The original AIML has some problems that must be resolved, namely its depth-first search function raised by [46, 47]. There are some simple ways to make the AIML search work very quickly. Standard AIML uses depth-first search, which does not optimise the result, as the name implies. It finds the first available solution by searching through a tree of possibilities.

In the proposed approach, the PMCBR algorithm uses the relational database to store a binary prefix tree of all the pattern strings listed in the AIML files. A recursive function

written in Perl searches the tree for a string that matches the input query. Once found, it dynamically retrieves the corresponding template from a second table. The template is stored as a single string containing AIML tags, which the Perl code must then parse. A relational database is used as a rapid way of accessing the information usually stored in AIML files. With enhanced version of AIML, the output sentence can be dynamically constructed using parts of the input sentence; therefore the relational database does not have to store every possible response.

### **3.8 Summary**

This chapter has discussed many existing technologies that could be combined into a CA framework for the Web, mobile services and CMC applications; these technologies are already in existence, but not as an integrated entity. From the experiment and proxy conversation, it can be anticipated that AINI's framework could play an important role in popularising the concept of intelligent CAs, not only on the web but also on mobile services or other network applications. This research paves the way for more humanoid user interfaces based on human language technologies.

It was also found that multilevel NL-Query models, DKMM, and modified *N*-tiered architecture address the extensive variability that is encountered in today's CA frameworks. Over-generalised interfaces make the framework hard to understand and use. There is a delicate balance between flexibility and simplicity. Performance cannot be compromised for the sake of flexibility and the lowest common denominator solution is often unacceptable. It is necessary to have flexible development environments, tools, and regression tests. Application interoperability is an essential direction for future improvement.

The AINI's framework has demonstrated that it could be deployed within any existing website to achieve intuitive user interaction for that website. The aim is to provide a framework that is easy to use and convenient for the developers. Personality and personalisation could be used within a CA's framework with no need for *redesign* and *remodeling*. The next chapter covers another important aspect on the assessment of the CA's knowledge bases extract from the Web. It addresses the issue of trustworthiness of knowledge bases and information extraction for CAs.

## CHAPTER 4

### AN ASSESSMENT OF THE TRUSTWORTHINESS OF KNOWLEDGE BASES FOR CONVERSATION AGENTS

#### 4.1 Introduction

This chapter explores the study of *agent knowledge* in the data server tier of the AINI's framework and particularly, its relationship to the related study of information trustworthiness and knowledge extraction. Some important methods for knowledge extraction from online documents are also discussed in this chapter. This is done in relation to the purpose of the design and development of domain-specific information for CAs' knowledge bases. This chapter focuses on a novel approach based on the proposed Web Knowledge Trust Model (WKTM) [173, 175, 190, 221-223] and the Automated Knowledge Extraction Agent (AKEA) [111, 112, 156]. The work of these proposals has been reported as research finding from this study and they are discussed in Sections 4.5 and 4.6.

To reiterate the purpose of this thesis, it aims to present a methodology to facilitate the use of online documents from the World Wide Web (WWW) and to use the acquired knowledge to develop the intelligent CA, AINI. Information extracted from public web pages could be problematic. There are issues that web pages may contain incorrect information or are outright hoaxes. Therefore, a WKTM is proposed in this study to determine 'trustworthy' websites, and, to ensure the credibility and reliability of the knowledge extracted from the web-derived corpora.

This study has applied WKTM to extract pandemic Bird Flu domain knowledge from the web using AKEA. WKTM is extensible through polymorphic inheritance of

components, and is based on a modular design. This chapter also provides an integrative review of the literature on the subject of websites' information trustworthiness. In addition, the WKTM applies both quantitative and qualitative approaches to determine the trustworthy websites. Within the context of this research, the term *authority* implies 'reputation', 'empowerment' and 'credentials', and the term *credibility* implies 'reliability', 'truthfulness', and 'unbiased' as characteristics of the selected websites.

At present, the WWW provides a distributed hypermedia interface to a vast amount of information. For instance, Google [224] currently has a training corpus of more than one trillion words (1,024,908,267,229) from public web pages. While the Web provides a huge source of information and data, commercial search engines however are not the best way to gather answers to queries due to the overwhelming number of results returned from a search. Nevertheless, despite certain obvious drawbacks such as the lack of control, there is no doubt that the WWW is a source of data of unprecedented richness and accessibility [225].

As reported in previous papers [113, 173] and discussed in Chapter 3, AINI's conversation agent operation is based on open-domain and domain-specific knowledge bases. Domain-specific knowledge bases consist of Natural Language Corpora and answers for Frequently Asked Questions (FAQ). Both components have been extracted from online documents using an Automated Knowledge Extraction Agent (AKEA) [156]. The AINI software agent was programmed to provide up-to-date information and to deliver essential information from trusted sources. The goal is that AINI will be capable of interacting with its users naturally and to provide reliable information.

The aim of this chapter is to develop intelligent agent techniques that will help to acquire information from websites that are reputable, credible, reliable and accountable.

The WKTM was developed with the specific aims as follows:

- to determine, through corpus analysis, the effectiveness of creating a CA's knowledge base with an unbiased corpus. The evaluation was based on data extracted from freely available online documents from the World Wide Web.
- to understand how WKTM can improve the selection of 'trustworthy' websites and most importantly, how this model can be applied to other domains.

## **4.2 Trust and Methodology**

There are 936 million pages<sup>66</sup> or between thirty to fifty thousand health-related sites available on the Internet. It has also been estimated that over 21 million users have been flooded by health information provided therein [226]. The Rhode Island HEALTH Web Query System [227] for example shown an ever-increasing public expectation and they are seeking for reliable and trustworthy online health information. The Rhode Island system also provides health professionals and community organisations with valuable information. However, less than half of the medical information available online has been reviewed by doctors [226]. There have been numerous detailed assessments of the quality of health information on the web. Eysenbach G. et al. [228] carried out a systematic review of health website evaluations and noted that the most frequently used quality criteria included accuracy, completeness, readability and the design of the site. There is a real need to explore and understand the ways in which the trustworthiness of how online health information are assessed. Unfortunately, few studies and little empirical evidence for a trust model in health-related websites exist [229, 230].

---

<sup>66</sup> Received 936 million pages hits by query on the Google search engine with the keyword "health" on 6 July 2007, <http://www.google.com/search?source=ig&hl=en&q=health>

**Table 4.1: Comparison of Trust Types and Measures in Previous Research<sup>67</sup>**

Study	Type of Trust	Aim of the Study	Experiment setup	Object of trust
McAllister (1995) [231]	<b><i>Cognition-based trust:</i></b> trust and respect people who are trustworthy because of good reasons and evidence.	This study addresses the relationships of interpersonal trust among managers and professionals in an organisation.	<b><i>Qualitative study:</i></b> Empirical study	Interpersonal trust in organisational settings
Parkhe(1998) [232]	<b><i>Institutional-based trust:</i></b> based upon intermediary mechanisms and by providing implicit guarantees	This research shows how partners can proactively manage an alliance relationship in order to develop trust.	<b><i>Qualitative study:</i></b> Empirical study	Inter-organisational trust
de Ruyter et al. (2001) [233]	<b><i>e-services -based trust:</i></b> exploring the antecedents of trust, relative advantage and perceived risk in the adoption of e-service	This research aims to investigate the impact of organisational reputation, relative advantage, and perceived risk on perceived service quality, trust and behavioural intentions of customers towards adopting e-services.	<b><i>Qualitative study:</i></b> Experimental study, participants were presented with offline role-playing scenarios	Customer trust
Kanawattanac hai & Yoo (2002) [234]	<b><i>Affect-based trust:</i></b> involves the emotional elements and social skills of trustees.	This research examines the differences of virtual teams in the changing patterns of trust over time.	<b><i>Qualitative study:</i></b> Empirical study	Interpersonal trust in virtual teams
Kim & Prabhakar (2002) [235]	<b><i>e-banking-based trust:</i></b> exploring initial trust in the adoption of online banking	This research hypothesized the consumers' propensity to trust, structural assurances to be antecedents of consumers' initial trust in the electronic channel as banking medium.	<b><i>Qualitative study:</i></b> Online survey	Online banking consumers trust
Gefen et al. (2003) [236]	<b><i>e-commerce-based trust:</i></b> exploring trust in an e-commerce vendor, using online travel agency	This research highlights that online consumer trust is as important as the Technology Acceptance Model (TAM) use-antecedents	<b><i>Qualitative study:</i></b> Experiential survey approach, participants performed product search at an online bookstore	Consumers trust online stores

<sup>67</sup> These empirical studies were collected from proceedings, journal papers and articles on the Internet. Some of the finding also available from Grabner-Kräuter & Kaluscha [236].

Study	Type of Trust	Aim of the Study	Experiment setup	Object of trust
Grabner-Kräuter & Kaluscha (2003)[237]	<b><i>e-commerce-based trust:</i></b> facilitating a multi-level and multi-dimensional analysis of research problems related to trust in e-commerce.	This research provides an integrative review of the empirical literature on trust in e-commerce in order to allow cumulative analysis of results.	<b><i>Qualitative study:</i></b> Online survey	Institutional phenomena (system trust) and personal and interpersonal forms of trust.
Paul & McDaniel Jr. (2004) [238]	<b><i>Relational-based trust:</i></b> is the extent one feels a personal attachment to the other party and believes that other party will intend doing good to him/her.	This research examines the relationship between interpersonal trust and virtual collaborative relationship performance.	<b><i>Qualitative study:</i></b> Field study	Interpersonal trust in virtual organisational settings
Leimeister et al. (2005) [239]	<b><i>System-based trust:</i></b> this type of trust is based on the perceived property of or reliance on a system or institution.	This study describes how trust enabling functionalities can be systematically designed and implemented in a virtual community	<b><i>Qualitative study:</i></b> survey	Interpersonal and system trust in virtual community
Ratnasingam (2005) [240]	<b><i>Technological -based trust:</i></b> confidentiality mechanisms, authentication mechanisms, and access controls mechanisms.	This research examines how institutional structures lead to technology trust for online B2B e-commerce.	<b><i>Qualitative study:</i></b> interview	Inter-organizational trust
Hsu et al. (2007) [241]	<b><i>Information-based trust:</i></b> exploring the nature of trust and divides it into three constructs—economy-based, Information-based and identification-based trust.	This research identifies the antecedents that support or hinder an individual's knowledge sharing behaviour by applying Social Cognitive Theory-based (SCT) model from both social environment and personal cognition aspects.	<b><i>Qualitative study:</i></b> Web-based survey	Virtual Communities (VC) trust



**Table 4.2: Website Trust Type and Methodology in Previous and Current Research**

Study	Type of trust websites	Aim of the Study	Methodology		Object of trust
			Theoretical	Experiment	
Koufaris & Hampton-Sosa (2002) [242]	<b><i>e-vendor:</i></b> exploring the antecedents of initial trust in an online company, using several e-vendors	This research provides customers with enjoyment and perceived control, which leads to greater trust in the company itself through the customers' perceptions about the web site's usefulness and ease of use.	Technology Acceptance Model (TAM)	<b><i>Qualitative study:</i></b> Experiential survey with on-line questionnaire, participants visited an Unfamiliar Web-site and performed a product search	Company websites
Fox & Rainie (2002) [229]	<b><i>e-health:</i></b> verifying online information is that health seekers trust the online environment.	This research is identifying a trusted online source for credible, comprehensive, and clinical healthcare information, and secure, confidential communications	-	<b><i>Qualitative study:</i></b> telephone interviews	Health websites
Fogg, Marable et al. (2002) [243]	<b><i>e-commerce, finance, health, news, search engines, sport, travel:</i></b> combine assessments of both trustworthiness and expertise to arrive at a final credibility perception.	This research aims to investigate what causes people to believe – or not believe – what they find online	Web Credibility	<b><i>Qualitative study:</i></b> comments of the websites	e-commerce, entertainment, finance, health, news, search engines, sports, and travel Websites
Corritorea et al. (2003) [244]	<b><i>e-commerce:</i></b> identifies three perceptual factors that impact on-line trust including perception of credibility, ease of use and risk.	This research proposes an online trust model, specifically trust between people and informational or transactional websites.	Online Trust Model	<b><i>Qualitative study:</i></b> interaction between users and websites	e-commerce Websites

Study	Type of trust websites	Aim of the Study	Methodology		Object of trust
			Theoretical	Experiment	
Sillence et al. (2004) [230]	<b>e-health:</b> identifying a design and information content factors influence trust and mistrust of online health sites.	This research aims to fill the gap in our knowledge of the ways in which real-world consumers evaluate information and online health advice (menopause).	Staged Trust Model	<b>Qualitative study:</b> observation and discussion	Health websites
Singer (2004) [245]	<b>e-business:</b> producing a more advanced and trust-based user interface and experience.	This research collaboration between MIT's Sloan Center for eBusiness and Intel aims to create a "trusted advisor" techniques and an online persona to help customers with a specific task on the Intel's Web-site.	WebTrust Methodology	<b>Qualitative study::</b> user click streams and user surveys	Intel's website
Goh & Fung [190, 222]	<b>e-health:</b> facilitating the use of trustworthy online documents relating to pandemic Bird Flu from the World Wide Web to create knowledge bases for intelligent conversation agents' (or bots')	This research proposes a Web Knowledge Trust Model to find 'trustworthy' websites and to ensure credibility and reliability of knowledge extraction based on web-derived corpora.	Web Knowledge Trust Model (WKTM)	<b>Qualitative study:</b> Computational linguistic  <b>Quantitative study:</b> PageRank and survey	Pandemic websites

Researchers in this discipline are also tackling the question on how to evaluate the quality and trustworthiness of online resources [243, 246-250]. Table 4.1 and Table 4.2 summarised an overview of the empirical studies on trust and the methodology. Pew Internet and American Life Project's Report [229] found that about a third of the Pew respondents felt the need to check the accuracy and reliability of the information they read.

As website and Internet technologies become more established, attention is turning to the factors that impact upon the success of websites. The key among these factors is "*trust*". Many studies have recognised that trust is a multidimensional construct and the studies have also examined different types of trust. Most of the studies were conducted in organisational settings or in the regime electronic commerce [231, 232, 236-238, 240, 244, 251]. There are different approaches to the study of trust across domains such as banking [235], management [231, 232, 234, 238], business [233, 242, 245], information and technology [239-241]; and others. However, although trust has been studied in a variety of disciplines, each of these disciplines has produced its own concepts, models, methodologies, definitions and findings. In fact, even within a given field, there is often a lack of agreement and focus of effort. Despite the nature or approaches in research on the subject of trust, researchers from every discipline do acknowledge the value of trust. Moreover, many scholars in HCI have begun to study trust in online information. The aims of such studies have ranged from looking at issues on design and interface elements, perceived website credibility [252, 253], and, the extent to which the technology is perceived and responded.

From a methodological viewpoint, this proposed research is based on a combination of qualitative and quantitative approaches. In the qualitative approach, an empirical survey was carried out by collecting comments from experts on web credibility, and on the other hand, the quantitative approach is based on log likelihood and Google's PageRank results from the

corpora collected from the web. According to literature [254, 255] a combination of methodologies used in research can potentially explore more advanced facets of online trust. Based on the result, both qualitative and quantitative research methods have made important contributions to the results in this research.

### **4.3 Websites as Objects of Trust**

There are many related concepts that are often confused with trust. Computer Mediated Communication (CMC) researchers study individual-to-individual trust relationships mediated through technology [256]. In contrast, other researchers focus on technology as the object of trust [257]. Researchers in the field of intelligent agents for example have looked at trust between software agents, a relationship in which agents can be objects of trust, and [258, 259] [11] have examined how people treat the intelligent agent as real people, and by extension, as objects of trust. They found that people do enter into “*relationships*” with computers, websites, companies, people and other new media. In this research, websites are defined as objects of trust in the WKTm. The term ‘website’ can be used to include the underlying information from online documents, the interactive user experience with the website, and/or the ownership of the website. In order to limit the scope of our study, this chapter will focus on the pandemic bird flu health related websites as the object of trust.

### **4.4 Trust Model**

There are few examples of trust model available for e-commerce, e-business, e-banking, e-health, etc as detailed in Table 4.2. Researchers at MIT's Sloan Center for e-Business and Intel, for instance, are trying to improve e-commerce trust with a new online system called WebTrust methodology [245]. The WebTrust methodology has enabled Intel to generate superior trust between Intel's Web site and customers. Koufaris M. and Hampton-Sosa W. [242] hypothesised the variables are perceived usefulness and perceived ease of use of the

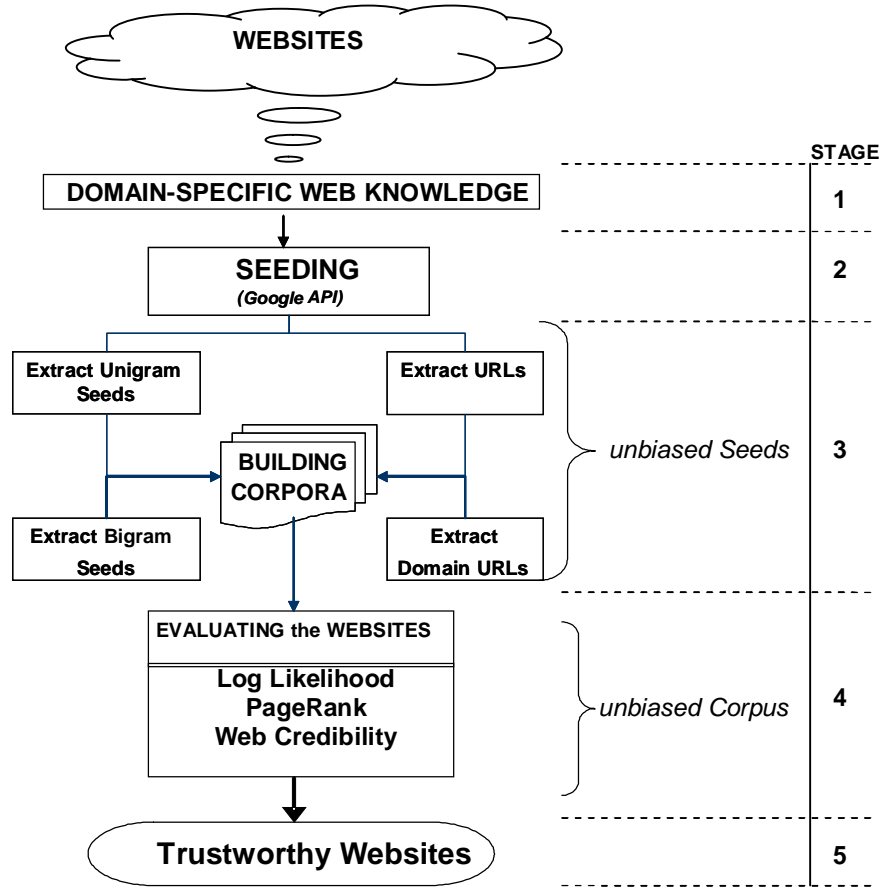
website based on Technology Acceptance Model (TAM) developed by Davis [260]. These two factors could be considered as the predictors for consumers' trust in the online company when they have their first contact with the Website. Cynthia. et al. [244], in proposing Online Trust Model, focused specifically upon trust between people and informational or transactional websites. In this model, they identified three perceptual factors that impact online trust including perception of credibility, ease of use and risk. The Staged Trust Model developed by Sillence et al. [230] focused their research in the field of online health advice, specifically for users seeking information on menopause. According to Sillence et. al., the information content on the website is the main factor influencing the trust and mistrust of online health sites.

Taking a closer look at the methodology and objects of trust used by other scholars, it was revealed that six of the sixteen studies (as shown in Table 4.2) focused on websites as objects of trust, and only three of those six examined health-related websites [229, 230, 243]. From these studies, two were conducted using the same research models (Technology Acceptance Model) [236, 242], but none has focused on how to select and extract online trustworthy websites for CAs. The following section describes in details the proposed WKTM approach.

#### **4.5 Web Knowledge Trust Model (WKTM)**

The objective of the WKTM is to provide solutions that will empower developers to adhere to the procedure described in Figure 4.1. It is expected that the model is also applicable to other application domains. The procedure outlined below is set out to address the question of *“how to select the trustworthy domain knowledge from existing online web documents?”* The WKTM procedure can be divided into five stages. First, the target of the web domain knowledge to be extracted is determined. For this study, pandemic Bird Flu is the focus of

the domain knowledge. In the second stage, a number of seeds are used in an iterative algorithm to bootstrap the corpora using unigram terms from the web.



**Figure 4.1: Web Knowledge Trust Model (WKTm)**

The process then proceeds to the third stage to extract bigram terms based on the final corpus and unigram terms extracted in the previous phase. Once the sets of domain URLs have been collected, they are then submitted as queries to the search engine via Google API (Application Program Interface)<sup>68</sup>. All the downloaded URLs will be used to build a final domain corpus. In the fourth stage, the corpus obtained are evaluated using Log Likelihood,

<sup>68</sup><http://www.google.com/apis>

Google's PageRank algorithm[261] and Stanford's Web Credibility criteria [243]. Finally, the top five most trustworthy websites will be selected and information will be extracted using AKEA.

#### **4.5.1 Selecting Domain-Specific Web Knowledge**

There are several sources from which web knowledge can be extracted; either from the well-edited and carefully balanced corpus such as the British National Corpus (BNC), or noisy but useful web documents which are largely unexplored[262]. The Google search engine claims to have indexed more than 14 billion pages on the web to date. Various studies [263], [264], [265] have indicated that, historically<sup>69</sup>, the web has doubled in size every nine to twelve months. Web pages are changing rapidly. In the period between 1<sup>st</sup> September 2006 and 20<sup>th</sup> October 2006, 3.36 billion pages were indexed by Google. This equates to about 67 million web pages added daily. With this huge volume of information, it is relatively "easy" to find a list of pages containing any given query terms. The difficult part is then to select, from the myriad of possible matching pages, the "top" 10 or 20 according to some computable quality measure which, ideally, closely resemble the user's notion of relevance. It is obvious that the ability of any search engine to closely match this human notion has a major impact on its success.

In this experiment, the Bird Flu pandemic is the focus of the domain knowledge base. In current times, pandemic flu has become an important research topic in order to meet the real-world challenge which may have severe global consequences. The Head of philanthropy at Google, Larry Brilliant, has also described his vision on how information technology can be used to fight pandemics [266]. However, as the Web becomes increasingly chaotic and has strong possibility of misleading and inaccurate health information, the Web could become

---

<sup>69</sup>Previous statistic of the Google Indexed pages can be access at [http://web.archive.org/web/\\*/http://www.google.com](http://web.archive.org/web/*/http://www.google.com).

harmful to the unwary users. Selection of trustworthy web pages is therefore an important factor in ensuring the long-term viability of the Web as a useful global information repository. The detailed descriptions of the subsequent stages in the WKTM are now given in the following sections.

#### 4.5.2 Seeding

The purpose of this stage is to select the corpus as a data acquisition resource for building the CA's knowledge bases. The objective of this stage is to create a "balanced" corpus of Web pages which contains relevant key words and documents of a given domain. For the purpose of seeding, we use words from the general training corpus, British National Corpus, (BNC)<sup>70</sup>. The BNC corpus consists of a collection of 100 million words in the forms of written and spoken language from a wide range of sources. It is designed to represent a wide cross-section of British English from the later part of the 20<sup>th</sup> century in both spoken and written forms. Since this research focuses on the Bird Flu pandemic, the initial seeds should come from its generic term derived from "bird" and "flu". From these seeds, we made a query to the online "specialised terminology" lists from the health information website MedLinePlus<sup>71</sup> Medical Dictionary. The term "bird flu" was found to be related to "avian influenza". With these four seed words, a query was sent to the BNC online corpus and the word "virus" was obtained as an additional seed. From the bigrams observation, the seed "virus" occurred 19 times in "flu virus" and 11 times in "influenza virus". Finally, we collected the five terms: "*bird*", "*flu*", "*avian*", "*influenza*" and "*virus*" for use as initial seeds for the investigation.

Once the seeds have been obtained, a comparison is made between the BNC corpus and Google's large-scale corpus from public Web pages. The purpose of the comparison is to

---

<sup>70</sup><http://www.natcorp.ox.ac.uk>

<sup>71</sup><http://www.nlm.nih.gov/medlineplus/plusdictionary.html>



determine whether the BNC corpus is covering similar terms or updated information as in the web. A comparison of the results from the two sources is shown in Table 4.3.

**Table 4.3: Comparing number of hit results from BNC and Google's Corpora using the set of unigram and bigram seed words**

<b>SEEDS</b>	<b>BNC</b>		<b>Google</b>	
<b>Unigram</b>	<b>Freq of BNC Counts</b>	<b>%</b>	<b>Freq of Web counts in '000s</b>	<b>%</b>
bird	3869	63.14%	14,400	33.13%
flu	573	9.35%	4,790	11.02%
avian	45	0.73%	1,360	3.13%
influenza	145	2.37%	2,120	4.88%
virus	1496	24.41%	20,800	47.85%
<b>Bigram</b>				
bird flu	1	3.23%	602	46.45%
avian influenza	0	0.00%	180	13.89%
flu virus	19	61.29%	206	15.90%
influenza virus	11	35.48%	308	23.77%

In Table 4.3, the Freq of count is the number of returns from searching BNC corpus and Google. As expected, the counts are much larger from Google than from the BNC. As shown in Table 4.3, the frequency of the total web counts from Google is 7,093 times larger than the BNC counts in the case of the unigrams. As for the bigrams, the Google Web counts are 41,806 times larger. These data were collected on 12<sup>th</sup> December, 2007. This evaluation demonstrates that BNC is small in terms of the number of frequency counts due to a smaller corpus as compared to Google. In addition, it can also be observed that the distribution of the seeds in the unigrams and bigrams are not similar. For instance, “avian influenza” as a scientific term for “bird flu” is not included in the BNC; whereas in the Google corpus, this term accounts for 13.89% of the returns from the seed queries. In addition, the colloquial term “bird flu” only occurred at a frequency of 3.23% in the BNC whereas in the Google corpus, the same term occupied almost 50% of the returns. From this exercise, it can be assumed that Google takes into account of the continual increase in the page volumes and scale-up its corpus accordingly. On the other hand, BNC has not been able to keep up with

newer terms such as “avian influenza” as indicated in Table 4.3. This also proves that BNC is insufficient by itself to provide the most updated information on any domain as in this case. However, as an initial stage in establishing the seeds for further query, the BNC has its merit as a training corpus. On the other hand, the Google returned over 600 thousands of web counts in the case of the seed word “bird flu”. This again makes any attempt to extract all the relevant knowledge from all these pages impossible. This therefore leads to the need to establish a more refined corpus and in particular, to acquire knowledge from trustworthy sites. The process is described in the following section.

#### **4.5.3 Building a Corpus**

In this stage, a domain-specific corpus on pandemic Bird Flu is built using the “*crawling*” approach. According to Broder et al. [267], crawling typically starts from a set of “seeds”. In this case, the seeds are obtained from the previous stage and consist of the five terms “*bird*”, “*flu*”, “*avian*”, “*influenza*” and “*virus*”. The crawling process consists of (a) fetch a page, (b) parse the page to extract all linked URLs, (c) for all the URLs not fetched previously, repeat steps (a) to (c).

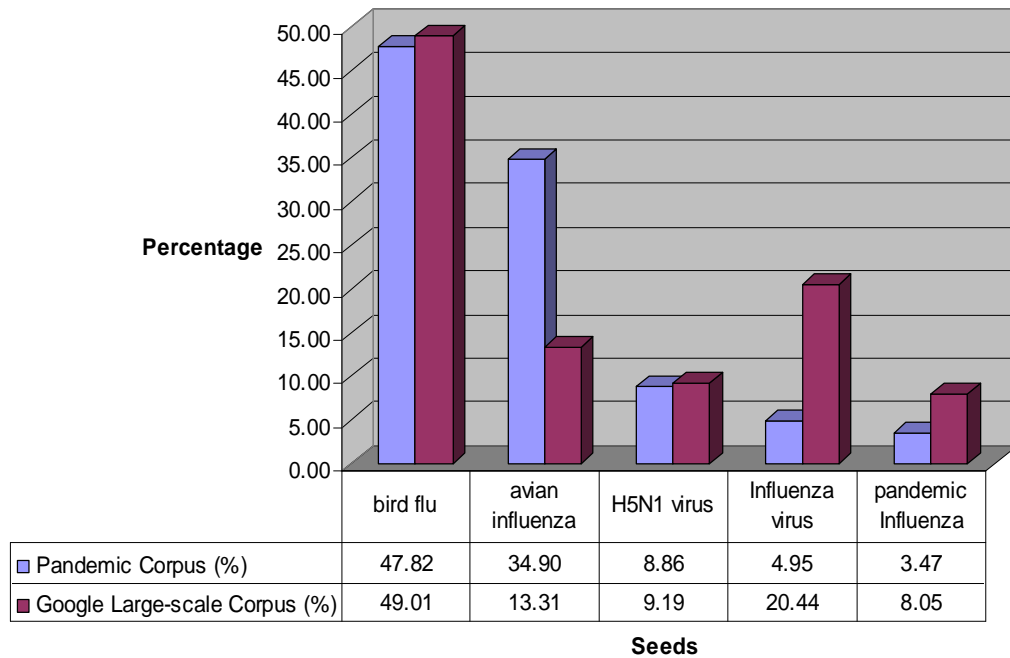
Normally, the crawling action will stop at some maximum value as limited by the Google API. For free service, Google limits the maximum number of queries to 1,000 per user per day. In this research, the number has been set as 10 URLs per search. The Google API is used to analyse the result rankings for several queries of different categories using statistical tools in the BootCAT Toolkit [268]. The corpora are essential resources for knowledge professionals who routinely work with specialised domain knowledge. BootCAT toolkit implements an iterative procedure to bootstrap specialised corpora and terms from the web and the process requires a list of “seeds” as input. The set of seeds are randomly combined, and each combination is used as a Google query string. These are the seeds which are

expected to represent the domain under investigation. A first query was sent to the Google search engine via Google API to extract the first corpus, and then new seeds were extracted from this corpus to build the final corpus [267].

Several important search parameters have to be controlled, such as the number of queries to be issued for each of the iteration, the number of seeds combined to build a query, and the number of pages to be retrieved for each query, and so forth. The first step of this phase is to extract a list of single- and two-word connectors from the corpus known as unigrams and bigrams respectively. During this phase, an additional seed called “H5N1” was found and it was frequently connected with other seeds in the final corpus. Hence, “H5N1” was added as the sixth seed to the seed set.

The next step is to retrieve the final URLs to build the final corpus. For simplicity and to avoid bias, only HTML and English pages are included. For each of the six seeds, BootCAT sends a query to obtain the number of URLs related to the seeds. In this process, the number of the final URLs returned is 1500 pages. After discarding the duplicated and broken URLs, the URL’s related to the domain under investigation is 1428.

A link analysis is applied to these sites under each domain name. If two domain names are linked with inbound and outbound connections, they are considered to be in a neighbourhood. Only the domains which are included in the neighbourhood are then selected. A few pages from each domain are then randomly chosen and concatenated into a document. After post-crawl cleaning, a corpus of 2,641,660 tokens is determined. This becomes the “Pandemic Corpus” in this research.



**Figure 4.2: Comparing distribution of seed words between the smaller set data Pandemic Corpus with the Google Large-scale Corpus**

In order to verify the usability of this established corpus, the distribution of returns was compared with respect to the larger Google corpus. This is shown in Figure 4.2. Although this corpus was created using a smaller set of seeds, it has a similar distribution as Google as seen from the figure. Hence it suggests that the unbiased method as described in this proposal yields a similar coverage proportion as Google. This leads to the next stage of evaluating the selected corpus and towards establishing the trusted and reliable domain knowledge bases.

#### 4.5.4 Evaluating a Corpus

Before one attempts to carry out an evaluation, it is necessary to define the term ‘trustworthiness’ associated with websites based on the credibility reports by [243] and [253]. Trustworthiness, a key element in the credibility calculus, is defined by the terms ‘reliable’, ‘truthful’, ‘unbiased’, and so on. Authority, another dimension of trustworthiness, is defined

by terms such as ‘authorised’, ‘reputable’, ‘accredited’, ‘credentialed’ and ‘empowered’. The word “authority” often indicates a government or an educational institution controlling the contents of a site. The authority dimension of trustworthiness associates with reputable organisations. Combining these two dimensions, this suggests that highly trustworthy websites will be perceived to have high levels of credibility [243], [252], [269], [270], [271], and authority. Based on these premises, this research is aimed at selecting the specific elements of a website that would lead to its consideration as a ‘trustworthy’ website. The elements proposed are based on Log likelihood ratio, PageRank and Web Credibility. They are described as follows.

#### **4.5.4.1 Log Likelihood Ratio**

In order to verify that the smaller pandemic corpus extracted by the proposed model is compatible to the large Google Corpus, the Log likelihood (LL) ratio is used as a quantitative assessment. The LL ratio approach is a statistical method in which a ratio is used to illustrate the coverage probability and accuracy within the confidence interval for two corpora. The higher LL ratio value indicates similar coverage probability even with small sample sizes [272] [273] [274].

The method is fairly simple and straightforward to apply. Tokens in a collocation tend to occur together more often than one would expect by chance. Statistical measures of association can be performed to determine the likelihood of the tokens in an N-gram occurring together more often than average. For example, if “bird” is continually followed by “flu”, the tokens “bird” and “flu” was defined closely associated with each other and they occur together more often than random. The statistical measures are computed using the various co-occurrence and individual frequency counts of an N-gram. For each word in the

two frequency lists, the LL statistics were calculated and constructed as a contingency table in Table 4.4:

**Table 4.4: Contingency table for word frequencies**

	Corpus 1	Corpus 2	Total
Frequency of word	a	b	a+b
Frequency of other words	c-a	d-b	c+d-a-b
Total	c	d	c+d

In the above table, the values ‘a’ and ‘b’ are called the observed values (O) from the two corpora. The values ‘c’ and ‘d’ corresponds to the number of words in Corpus 1 and 2 respectively. They are the two N values. The expected values (E) are calculated based on equation (4.1):

$$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i} \quad (4.1)$$

where  $N_1 = c$ ,  $N_2 = d$ , word  $E_1 = c*(a+b) / (c+d)$  and  $E_2 = d*(a+b) / (c+d)$ .

The calculation for the expected values takes account of the size of the two corpora. Based on this expected value, the LL ratio was calculated using equation (4.2).

$$-2 \ln I = 2 \sum_i O_i \ln \left( \frac{O_i}{E_i} \right) \quad (4.2)$$

In this case where  $i = 2$ , the calculation will be based on:

$$LL = 2*((a*\log (a/E_1)) + (b*\log (b/E_2)))$$

The word frequency list is then sorted by the resulting LL values. This gives the effect of placing the largest LL value at the top of the list representing the word which has the most significant relative frequency difference between the two corpora. In this study, the bigrams-

based version of the LL measure in the *Ngram* Statistical Package (NSP)<sup>72</sup> is used. In Table 4.5, the high LL-score values indicate the most important similarities between the two corpora for the coverage of the seed words. The results show that the proposed approach produces a confidence interval with a similar coverage probability and a high level of accuracy based on the seed words from the two corpora.

**Table 4.5. Log-likelihood Ratios for Pandemic Corpus vs Google large-scale Corpus**

Bigram	Pandemic Corpus	Google Large-scale Corpus in '000s	LL- Score
bird flu	12640	27,100	+106266.72
avian influenza	9223	7,360	+95698.31
H5N1 virus	2342	5,080	+19635.16
Influenza virus	1307	11,300,000	+ 7387.20
Pandemic Influenza	918	4,450	+ 6233.06
Total Corpus	2,641,660	1,024,908,267	

#### 4.5.4.2 PageRank

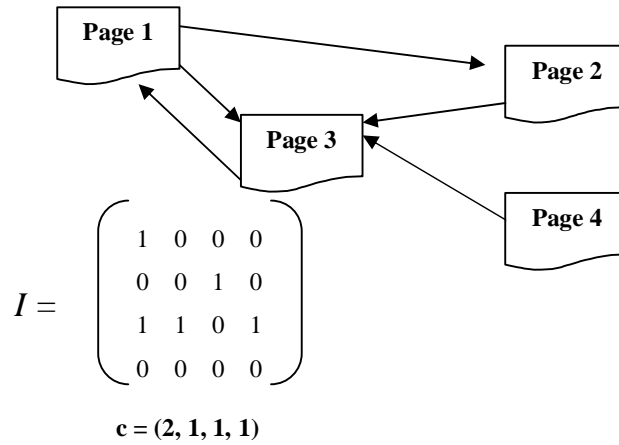
Evaluating enormous amounts of websites manually is not an easy task. Another approach is to use the Google's PageRank algorithm [261]. PageRank is a unique democratic process relies on the nature of the Web by using the web's vast link structure as an indicator of an individual page's value. It is the core algorithm of the Google's search engine. The algorithm is a complex and automated method which makes human tampering with the PageRank results extremely difficult. It should be noted that Google does not sell placements within the results thereby maintaining the democratic and unbiased nature of the search results. In this research, PageRank is used as one of the criteria to evaluate the trustworthiness of the websites based on link analysis. A similar application of link analysis is the evaluation of the quality of an academic work by analysing the amount of citations. The number of backlinks to a given page gives some approximation of a page's importance or quality. PageRank

<sup>72</sup> NSP Package can be downloaded at <http://search.cpan.org/~tpederse/Text-NSP-1.03/>

extends this idea by not considering the links from all pages as equal. The algorithm also normalises the final value to a range of 0 to 10. PageRank is defined in the equation 4.3:

$$p_i = (1-d) + d \sum_{j=1}^n (I_{ij} / c_j) p_j \quad (4.3)$$

Suppose  $n$  is the number of webpages. Let  $I_{ij} = 1$  if page  $j$  points to page  $i$ , and zero otherwise.  $c_j$  is the number of pages pointed to by page  $j$  (i.e., number of outlinks). The Google PageRanks  $p_i$  is defined by the recursive relationship where the parameter  $d$  is a damping factor which can be set between 0 and 1. In this study, the value of  $d$  is usually set to 0.85. Note that the PageRanks form a probability distribution over web pages, so the sum of all web pages' PageRanks will be one. PageRank or  $p_i$  can be calculated using a simple iterative algorithm, and corresponds to the principal eigenvector of the normalised link matrix of the web.



**Figure 4.3: An Example of PageRank Corresponding to Web pages**

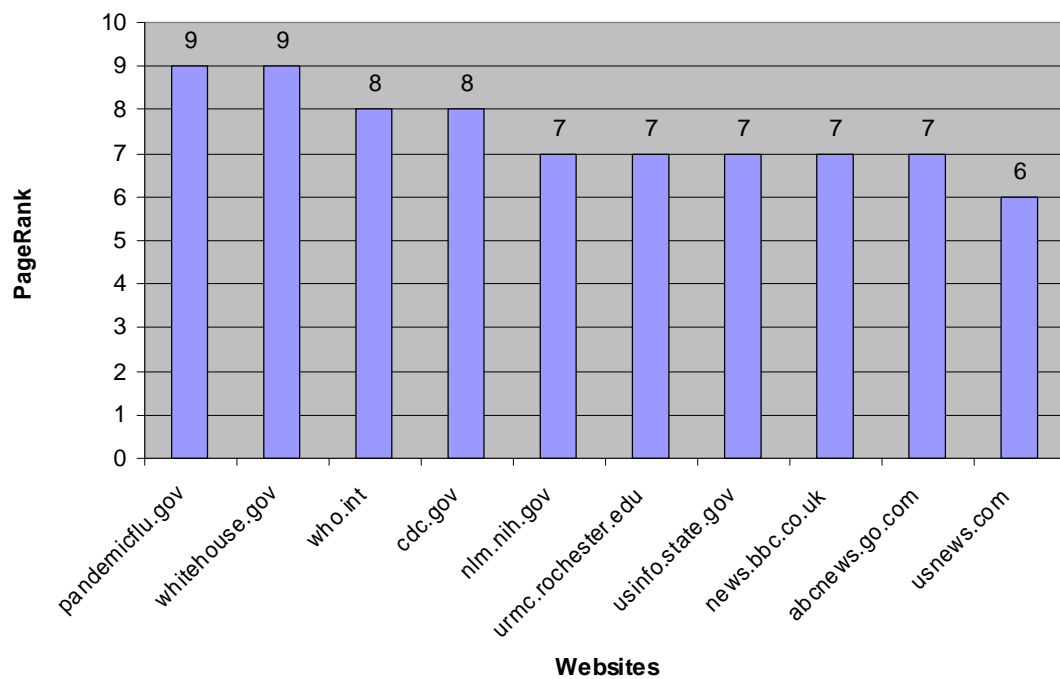
**Solution:**  $p_i = (1.49, .78, 1.58, .15)$

Note that Page 4 in Figure 4.3 shows no incoming links, and hence gets the minimum PageRank of 0.15. Google PageRank uses a whole number of 1 as a base for each page, and a 'voting value' of .85 damping factor of the PageRank value. (This being the result of



the .15 damping factor applied to all pages when determining the value of outbound links from a page).

Based on the 1,428 URLs returned from stage 3, a query is sent to Google's PageRank directory to determine their rankings using Google PageRank algorithm (see Appendix C). The algorithm will calculate the PageRank of a specified URL and returns an integer value from 0 to 10. Figure 4.4 shows the results of the top ten sites based on the PageRank scale<sup>73</sup>.



**Figure 4.4: Top Ten PageRank Scale for the Bird Flu Domain**

The least important site is one with a PageRank of 1. The most referenced and supposedly important sites are those with a  $P_i$  of between 6 and 10.

<sup>73</sup> The results of the PageRank extracted on 25 January 2007 and the results will be difference based on algorithm 4.3.

#### 4.5.4.3 Web Credibility

This section presents the credibility assessment of the top 10 websites related to this study assessed by a form of qualitative approach. After the PageRank results have been collected from the top 10 sites, a site is assigned with scores manually by experts based on the Web Credibility ranking criteria [243]. In this experiment, ten experts from the American Association of Webmasters<sup>74</sup> in the web design field were asked to assess the credibility of these sites based on their professional judgement.

Participants were given the opportunity to leave brief comments about the credibility of a site randomly assigned to them. The rankings are based on the 18 types of comments as shown in Table 4.6. The participants' assessment comments for each site's credibility were analysed. The comments were coded into two categories: *trust* or *distrust*. Some of the comments coded in this category are as shown in Table 4.7. The percentages shown in Table 4.6 represent how often a comment on that topic appeared in the entire set of comments. For instance, participants commented on the appearance and information design of the site more often than any other website feature, with these comments accounting for 39.9% and 37.6% (respectively) of the total of 168 comments.

**Table 4.6: Frequency of the Comment Topics for Website Credibility**

Comment Topics (addressing specific credibility issue)	Percentage ( of 168 comments)
Design Look	39.9%
Information Design/Structure	37.6%
Information Accuracy	29.2%
Information Bias	24.8%
Information Usefulness	21.6%
Information Focus	20.1%
Information Clarity	19.4%
Name Recognition and Reputation	17.1%
Company Motive	16.5%
Advertising	14.8%
Writing Tone	9.7%
Identity of Site Operator	8.4%

---

<sup>74</sup> <http://www.aawebmasters.com>

Site Functionality	7.9%
Past Experience with Site	5.3%
Performance on Test by User	4.1%
Readability	3.9%
Customer Service	3.7%
Affiliations	2.5%

The data shows that when participants evaluated the credibility of sites, they focused relatively more on issues of design (design look and information design) and trustworthiness (the accuracy, focus, usefulness, bias and clarity of information). Table 4.7 shows the sample of comments related to trustworthiness. As experts in web design, a majority of the participants ignored the issues that were not related to their expertise. Hence, most of the participants relied more heavily on other areas perceived to be more important, such as appearance and information design, accuracy and information dissemination [243].

**Table 4.7: An Example of Comments Related to Trustworthiness of the Website Information**

Topic	Trust (Positive Comments)	Distrust (Negative Comments)
<b>Information Focus</b>	<ul style="list-style-type: none"> <li>• <i>This site was committed to news and the dissemination of information.</i></li> <li>• <i>Well focused and well organized, which lends to more credibility</i></li> <li>• <i>The site looks credible because of the extensiveness of information.</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>This site was credible, but too wide-ranging the information.</i></li> <li>• <i>Information badly presented on the site and too much information in single page.</i></li> <li>• <i>This site seems focused on drugs and little on the health information.</i></li> </ul>
<b>Information Accuracy</b>	<ul style="list-style-type: none"> <li>• <i>This site provided useful and interesting knowledge or facts.</i></li> <li>• <i>Most of the articles on this Web site seem to be headline news and can be heard elsewhere.</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>Less credible and the fact that it is crap and just to take an attention from the users.</i></li> <li>• <i>The information displayed on the website harder to verify.</i></li> </ul>
<b>Information Bias</b>	<ul style="list-style-type: none"> <li>• <i>It is credible because the opinions contained therein are based on authority websites.</i></li> <li>• <i>The World Health Organisation has a great reputation. I would trust the information found in this website.</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>This site looks professional but based upon personal opinion and unaware of up-to-date information.</i></li> <li>• <i>This site is more commentary, and intolerant. The arguments are more debatable, and thus less "trusted."</i></li> </ul>
<b>Information Usefulness</b>	<ul style="list-style-type: none"> <li>• <i>This site provided useful and interesting knowledge about health information.</i></li> <li>• <i>I find this site useful and informative because it offers services to the community.</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>This site appears not very useful and more like other news websites.</i></li> <li>• <i>I feel the advertising on the site were off-putting and reduced the credibility of the site</i></li> </ul>

<b>Information Clarity</b>	<ul style="list-style-type: none"> <li>• <i>This site spells out very clearly their mission and activities. Information presented on the site is easy to understand and straight to the point.</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>I found this site too much text and the information puzzling.</i></li> <li>• <i>Nothing on the site but full of drugs terminology.</i></li> </ul>
----------------------------	---	---

Participants thought that the visual appeal of the site was important; poor visual appeal did not encourage further exploration. Poor design gave a negative first impression and the name of the website could lead participants to mistrust the site and its author's intentions. Participants were also influenced by the websites' names - 17.1% of comments related to *name recognition and reputation*. A good name was specific and to the point, but was not patronising or too gimmicky. A poor name was not trustworthy and could lead to a rapid rejection of the site. Most individuals preferred sites that are run by reputable organisations, or had a medical or expert feel about them. They trusted the information on such websites especially when the credentials of the site and its authors were made explicit. Sites that indicated that the information originated from a government organisation (.gov) or education entity (.edu) were well received.

Table 4.6 indicates that content or information factors were amongst important features in describing trusted or well-liked sites. Participants trusted the selected sites because they demonstrated an in-depth knowledge of a wide variety of relevant topics and put forward clear and unbiased information. Participants were more likely to trust the information if they could verify it and cross check it with other websites. Participants placed the most trust in sites that provided informative content on a wide range of relevant topics. The information was trusted if it was unbiased, and if the information on such sites was supported by research articles or original sources. Sites that were selected contained a variety of content features including Frequently Asked Questions (FAQs) and a section on hints and tips. This result is supported by Stanford's Web Credibility large study[243]. Their data suggests that participants evaluating the credibility of health websites pay relatively more attention to the

focus and usefulness of the information. What participants find in these areas apparently becomes a significant indicator of the site's credibility.

The 'Top 10' sites collected from Google PageRank were then ranked according to their mean scores, highest to lowest. This ranking gives a general idea about which sites in this study have been found to be the most or the least credible by the users. When a more credible site was listed on the page, the site's score was given a point and the less credible site lost a point. Over the course of the study, each site was evaluated many times, gaining and losing points along the way. At the end of the study, each site received a final score, which was the average (mean) of all the scores it had received from the experts. The average value is the total number of points divided by the total number of times the site was ranked. If a site has a score of +1.0, it means the site is deemed to be credible by all participants. If the score is 0.0, it means the site was considered to be credible half of the time. Combining the three methods described, Table 4.8 shows the results of the trustworthiness analysis for the top 10 sites related to the domain knowledge in this study.

**Table 4.8: Website Rankings based on Google's PageRank and Stanford's Web Credibility**

Website	URL	Web Credibility Average Score	PageRank Score	Total Score	Final Ranking
U.S. Department of Health & Human Services	<a href="http://www.pandemicflu.gov">pandemicflu.gov</a> <sup>75</sup>	0.56	9	9.56	1
The White House	<a href="http://www.whitehouse.gov/infocus/pandemicflu/">whitehouse.gov</a> <sup>76</sup>	0.42	9	9.42	2
World Health Organization	<a href="http://www.who.int/csr/disease/avian_influenza/en/">who.int</a> <sup>77</sup>	0.54	8	8.54	3
U.S. Centers for Disease Control and Prevention	<a href="http://www.cdc.gov/flu/avian/">cdc.gov</a> <sup>78</sup>	0.41	8	8.41	4
U.S. National Library of Medicine	<a href="http://www.nlm.nih.gov/medlineplus/birdflu.html">nlm.nih.gov</a> <sup>79</sup>	0.46	7	7.46	5
University of Rochester	<a href="http://www.urmc.rochester.edu">urmc.rochester.edu</a>	0.37	7	7.37	6

<sup>75</sup> <http://www.pandemicflu.gov/>

<sup>76</sup> <http://www.whitehouse.gov/infocus/pandemicflu/>

<sup>77</sup> [http://www.who.int/csr/disease/avian\\_influenza/en/](http://www.who.int/csr/disease/avian_influenza/en/)

<sup>78</sup> <http://www.cdc.gov/flu/avian/>

<sup>79</sup> <http://www.nlm.nih.gov/medlineplus/birdflu.html>

Medical Center	<sup>80</sup>				
U.S. Department of State's Bureau of International Information Programs	<a href="http://usinfo.state.gov/birdflu/">usinfo.state.gov</a> <sup>81</sup>	0.03	7	7.03	7
BBC News	<a href="http://news.bbc.co.uk/1/hi/in_depth/world/2005/bird_flu/default.stm">news.bbc.co.uk</a> <sup>82</sup>	-0.45	7	6.55	8
ABC News	<a href="http://abcnews.go.com/Health/AvianFlu/">abcnews.go.com</a> <sup>83</sup>	-0.54	7	6.46	9
U.S. News & World Report	<a href="http://www.usnews.com/usnews/health/bird-flu/">usnews.com</a> <sup>84</sup>	-0.29	6	5.71	10

What Credibility Average Scores Mean:

+0.50 = judged as more credible in 75% of pairings

0.0 = judged as more credible in 50% of pairings

-0.50 = judged as less credible in 75% of pairings

#### 4.5.5 Trustworthiness Websites

The final set of URLs was further culled to include only selected sites attributed to regulated authorities. They are mainly government bodies, international organizations or educational institutions. All these organizations control and provide the contents of their respective sites. Once the seed set is determined, each URL's page is further examined and rated as either reliable or reputable. As shown in Figure 4.5, the selection is reviewed, rated and tested for connectivity with the trusted seed pages. The expert participants in the web credibility assessment exercise preferred websites that contain a great deal of information, instead of publicity news from the media such as BBC News, ABC News and USNews. These results also showed that the content or information factors were more important than design features in describing trusted or well-liked sites. In the current study, the final five websites cluster at the top of the web trustworthiness rankings are: *pandemicflu.gov*, *whitehouse.gov*, *who.int*, *cdc.gov* and *nlm.nih.gov*. All these highly credible sites were selected based on PageRank and credibility scale scores. These five top sites are clearly viewed by the expert participants as more credible than the other five sites in this study.

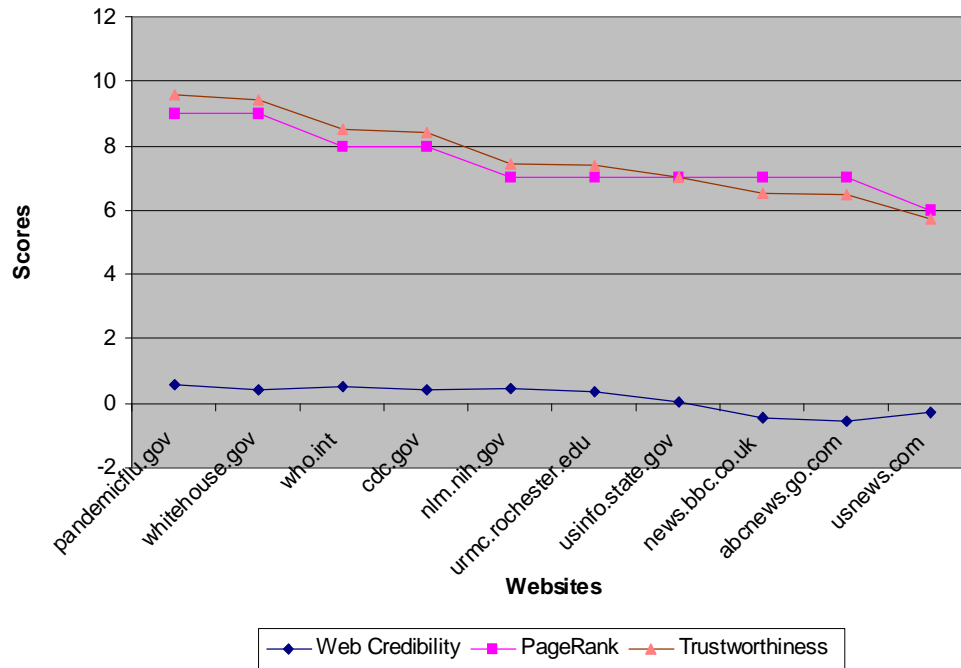
<sup>80</sup> [http://www.urmc.rochester.edu/pr/current\\_research/bird\\_flu/index.cfm](http://www.urmc.rochester.edu/pr/current_research/bird_flu/index.cfm)

<sup>81</sup> <http://usinfo.state.gov/birdflu/>

<sup>82</sup> [http://news.bbc.co.uk/1/hi/in\\_depth/world/2005/bird\\_flu/default.stm](http://news.bbc.co.uk/1/hi/in_depth/world/2005/bird_flu/default.stm)

<sup>83</sup> <http://abcnews.go.com/Health/AvianFlu/>

<sup>84</sup> <http://www.usnews.com/usnews/health/bird-flu/>



**Figure 4.5: Comparing Trustworthiness of Top 10 Websites related to the Bird Flu Domain**

The results support the proposal that the trustworthiness of websites is not only based on the PageRank and Web Credibility, but also the ‘authority’ of the websites which is not taken into account within the PageRank and Stanford Web Credibility criteria. There are other important factors in determining the ‘reliable authority’ of a site. They could be based on the site’s history and the number of back-links to government agencies, education institutions, and international organizations. The more established and relevantly linked a site has, the more likely it could be considered as ‘stronger’ or ‘more reliable’. This may effectively suggest the linked site has ‘authority’, ‘reputability’, ‘empowerment’ and ‘credentials’. This work will be examined in future study. Finally, the top five URLs are then used as the main source of knowledge for AKEA to extract the pandemic related contents to build AINI’s domain-specific knowledge base.

## 4.6 Automated Knowledge Extraction Agent (AKEA)

Automated Knowledge Extraction Agent (AKEA) was developed during the breakout of the Severe Acute Respiratory Syndrome, or SARS, epidemic in 2003. [111, 112, 117, 156]. As mentioned earlier, our objective is to deliver essential information from trusted sources. The AKEA module is intended to build a knowledge base for the CA automatically. The information will be extracted and mined from the existing top five selected trustworthy websites described in Section 4.5. Figure 4.6 shows a summary of AKEA's architecture. The framework consists of seven modules developed for different applications [212, 275].

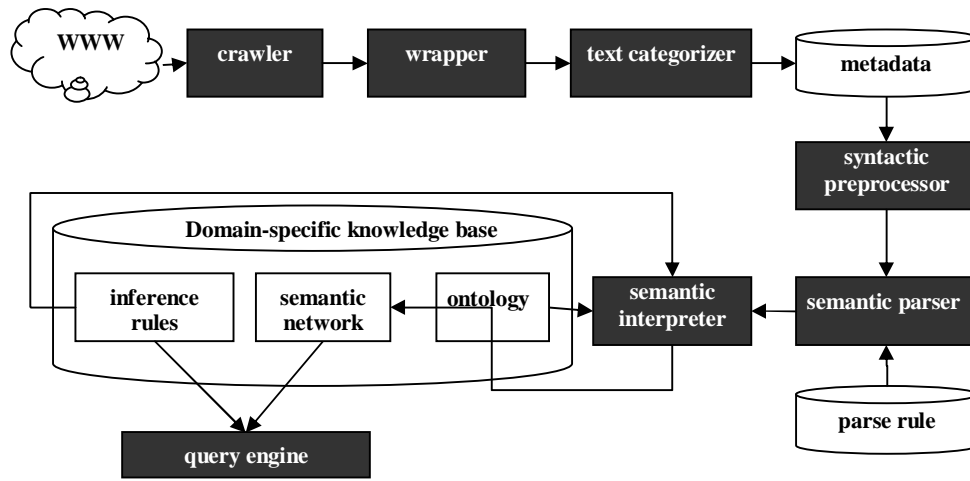


Figure 4.6: AKEA Architecture

### 4.6.1 Crawler

Given a selected trustworthy URL as a site of interest, the crawler will traverse the pages and follow the links in a breadth-first manner and will return all pages that meet requirements for further filtering. The functions of the crawler are like those used in conventional crawler-based search engines [261, 276]. The crawler resolves root domain selected from the trusted websites discussed in Section 4.5, ie. *pandemicflu.gov*, *whitehouse.gov*, *who.int*, *cdc.gov*, *nlm.nih.gov*, and follows subsequent links which are available on a page until a certain depth



as defined by the user and shown in Figure 4.7. These configurations are set in the crawler database. For every page crawled, a copy is returned for further processing by the wrapper. The activities of the crawler are logged in the crawler log database.

The screenshot displays a web interface titled "administration" with a progress indicator. It features two main sections: "Crawler List" and "Edit Crawler Link".

**Crawler List:** This section contains a list of crawler names: "U.S. Department of Health & Human Services", "The White House", "World Health Organization" (which is highlighted with a black border), "U.S. Centers for Disease Control and Prevention", and "U.S. National Library of Medicine". To the right of each name is a small icon representing a document with a magnifying glass.

**Edit Crawler Link:** This section is used to configure a selected crawler. It includes the following fields:

- Crawler Name:** A text input field containing "World Health Organization".
- URL:** A text input field containing "http://www.who.int/csr/disease/avian\_influenza/en/".
- Domain:** A text input field containing "http://www.who.int".
- Max Size:** A text input field containing "20000".
- Min Size:** A text input field containing "100".

A "Submit" button is located at the bottom of this section.

A large, curved arrow points from the "World Health Organization" entry in the "Crawler List" to the "Crawler Name" field in the "Edit Crawler Link" section, indicating the selection and editing process.

**Figure 4.7: Crawler's Configuration Interface**

#### 4.6.2 Wrapper

The wrapper will extract the title of the page, author, date, content and other metadata encoded in the page. One of the notable features of this framework is that it is open for the exploration of any possible XML-annotated ontological information. In other words, this module is extensible to include future semantic web facilities. All this information is useful as an additional semantic to describe the knowledge encoded in the pages and to expand the knowledge base.

#### **4.6.3 Text Categoriser**

The text categoriser reads the content returned by the wrapper and, using Naïve Bayes probabilistic classifiers [21], the categoriser will then assign category labels to these pages. The category information is stored in a buffer and it is important to conclude whether a page in the buffer should be processed any further. This acts as a secondary filter after the crawler to ensure that only the desired web documents are further processed for the CA's knowledge base.

#### **4.6.4 Syntactic Preprocessor**

This stage reads the categorised text, tokenises them and identifies the dependencies among them. Based on the dependencies, grammatical relations, that is, phrasal categories such as noun phrases, verb phrases and prepositional phrases, are extracted based on the formal grammatical system called X-MINIPAR. X-MINIPAR is a modified version of MINIPAR [211] (see details in Section 5.2). The named entities in noun phrases are tagged with identifiers such as organisation and person which will be used later as predicates for the first-order logic. Finally, the morphological root of verb phrases is identified for later use.

#### **4.6.5 Semantic Parser**

At this stage, the parser will translate the syntactically-tagged text and the relevant metadata into semantic representation in the form of first-order logic terms using certain parse rules. Morphological roots of verb phrases will form the factor for relations and noun phrases will be the arguments in the relevant relations or properties.

#### **4.6.6 Semantic Interpreter**

Using the definition of concept classes, relations, properties and inference rules, the interpreter tries to fit into the ontology by producing semantic instances explaining the meaning of the current text. These instances are then stored into the knowledge base as a semantic network.

#### **4.6.7 Query Engine**

The query engine allows the use of the domain-specific knowledge base to produce better retrieval or to perform inferences, and to a certain extent, to be used to assist in problem solving. There are three data structures that support the operation of the module as described below:

- **Metadata**

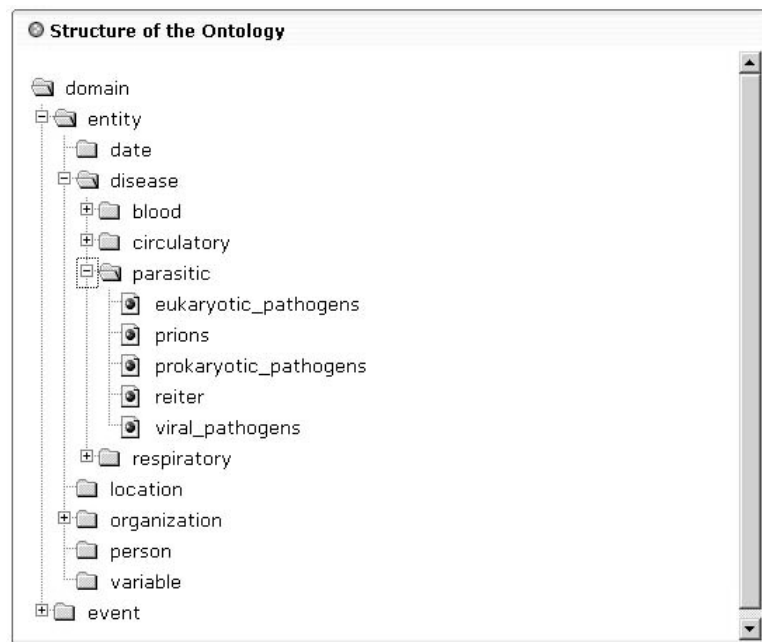
Metadata consists of information that can be inferred directly from the syntactic structure of HTML pages. This information is used to identify and define each page such as the title, author, content, data, keywords, category, etc.

- **Parse Rule**

The parse rule consists of rules stating how predicates and arguments are to be placed together to form logic representations.

- **Knowledge Base**

The knowledge base consists of three parts: semantic network, ontology and inference rules. The semantic network is in a form understandable and able to be manipulated by machines. The ontology is a conceptual schema that contains the definition of concepts of interest, relations between them and their properties. It can be seen as a template of classes where the instances of these classes will populate the knowledge base. As for the inference rules, they are used by the query engine to intelligently infer new and untold facts from existing ones.



To view this ontology on different page, click [here](#)

**Figure 4.8: Ontology Structure in AKEA**

As discussed in Section 3.5.3 on DKMM, AINI’s domain-specific knowledge bases are made up of two unit knowledge domains called Natural Language Corpus and FAQ component. The unit domains in the Natural Language Corpus component consist of knowledge and information harvested from or expressed in ontologies (see Figure 4.8), gazetteers and named entities (see Figure 4.9).

NAMED-ENTITY RECOGNITION FOR query (SENTENCE 0 )				
INPUT	GRAMMATICAL CATEGORIES & RELATIONSHIPS			
	1	N	wh	What fin (E0)
	2	VBE	i	are fin (E0)
	3	Det	det	the concern (4)
	4	N	pred	concern be (2)
	5	Prep	mod	about concern (4)
	6	A	mod	avian influenza (7)
	7	N	pcomp-n	influenza about (5)
	12	Det	det	A H5N1 outbreaks H5N1 (10)
	13	Prep	mod	in outbreak (12)
	14	N	pcomp-n	Asia in (13)
	16	N	conj	Europe outbreak (12)
	18	Det	det	the Near East (20)
	20	N	conj	Near East Europe (16)
	23	N	conj	Africa Near East (20)

**Figure 4.9: Name Entity Recognition in AKEA**

These have been implemented as domain-dependent modular components which will allow future improvements in the domain knowledge. The named entity module identifies named locations, named persons, named organisations, dates, times and key measures in text formats. The information is obtained by AKEA. In order to identify these entities, our system uses rules to specify the named entities' structure in terms of text tokens and information from the source such as tagger, morphosyntactic analyser and knowledge bases of names, clue words and abbreviations. Figure 4.9 shows an example of the name entity recognition for the sentence “*What are the concerns about avian influenza A (H5N1) outbreaks in Asia, Europe, the Near East, and Africa?*.”<sup>85</sup>

## 4.7 Summary

This chapter has discussed how WKTm can be used to evaluate the trustworthiness of websites as the objects of trust for the development of CAs' knowledge bases. This chapter also resolved many of the issues raised from the first full implementation of AKEA during the SARS crisis in 2003 [110-112], in which the knowledge extraction was based on the popularity of websites, instead of their trustworthiness.

Based on the proposal and experiment described in this paper, the contributions of this chapter are:

- The procedure of selecting trustworthy websites for building a conversation agent's knowledge bases is proposed.
- A scheme for selecting a “unbiased seed set” for building a corpus has been presented.
- A Web Knowledge Trust Model (WKTm) for determining reputable, credible, reliable and accountable websites is proposed.
- Results of an evaluation based on 1,428 Bird Flu Pandemic websites crawled by Google API are presented and discussed. Some interesting statistics on the hit

---

<sup>85</sup> The example of the sentence extracted from <http://www.pandemicflu.gov/faq/avianinfluenza/1235.html>

frequency, a significant data collection based on PageRank and Stanford Web Credibility are observed. The corpus is also used to evaluate the proposed WKTM.

These contributions indicate that this novel approach contributes towards the building of restricted CAs domain knowledge based on WKTM. The proposed model demonstrates the credibility of the web sites could be defined and is probably closer to a realistic expectation of trustworthiness. The course of this research is now directed to the evaluation of the CAs framework. The practical development, implementation and testing of the framework are dealt with in the following chapters.

## **CHAPTER 5**

### **AN EVALUATION OF CONVERSATION AGENTS FRAMEWORK**

#### **5.1 Introduction**

The growth of conversation systems is limited by a lack of evaluation, especially of the performance and the quality of responses from the CA's. The Turing Test, for instance, is a proposal to test a machine's capability to demonstrate intelligence in five minutes. In the Loebner Prize contest, judges evaluate which of the CAs entered is the most humanlike in ten minutes. For a practical CA operating without the time limit, a novel methodology based on qualitative and qualitative approaches is proposed. It is apparent that in evaluating a CA framework, quantitative means such as statistics and facts are required to decide whether a development is a step forward, and, whether the development is worth the effort.

In this research, the conventional approach based on laboratory experiments is used. This is followed by a real-time experiment and the results are compared to previous empirical research. These aspects of work are described in Chapter 6 and Chapter 7 respectively. In addition, mixtures of different methods (qualitative and qualitative) are used in the evaluation. They are described in the following sections.

#### **5.2 An Evaluation of the Parsers**

As discussed in the Section 3.5.4.3, off-shelf modified version of MINIPAR [211] called X-MINIPAR was used as the parser in the AINI's NL-Query module. A comparison of X-MINIPAR was made with other popular parsers. In this study, X-MINIPAR was compared to

the CMU Link Grammar parser [277] and the Stanford parser [278]. These three parsers represent a cross-section of approaches to producing dependency analyses: X-MINIPAR uses a constituency grammar internally before converting the result to a dependency tree. CMU Link Grammar is based on link grammar, and the Stanford Parser is an unlexicalised statistical syntactic parser. With the emergence of broad-coverage parsers, quantitative evaluation of parsers becomes increasingly more important. Firstly, such an evaluation scheme is necessary to quantitatively measure the progress in the field of broad-coverage parsing and to compare and evaluate different parsing techniques. Secondly, in the development of a broad-coverage parser, it is usually very difficult to predict the consequences of a change to the parser or the grammar. An attempt to extend the coverage that is motivated by a few examples may well cause the parser to over generate or lose coverage in other areas. This will make the evaluation biased and difficult to compare. Thirdly, efficiency and coverage are often conflicting goals for a parser. A meaningful trade-off can only be arrived at if both of them can be measured precisely. Finally, quantitative evaluation may provide crucial information for determining the suitability of a parser in a particular domain or for a particular task.

In this evaluation process, three parsers are installed and an extrinsic evaluation is applied as described by Bangalore et al. [279]. Extrinsic evaluation is usually used as an indirect method for comparing parsing systems. This approach is acceptable even if the parsers produce different representations for their outputs, as long as the output can be converted into a form usable by the system. The parsers are evaluated in terms of performance and accuracy of the output.



### 5.2.1 Performance of the Parsers

In order to evaluate the comparative performance of the parsers, 1428 uncategorised pandemic bird flu web documents are extracted and wrapped using Google API<sup>86</sup> and BootCAT Toolkit [268]. The process has been detailed in Section 4.5.3. Regular expressions are used to filter out the HTML information and to extract well-formed sentences. From the 1428 pages, the first sentence of each of the first 150 pages are extracted and parsed by X-MINIPAR, CMU Link Grammar and the Stanford parser. Each sentence has an average of 50 words. A uniform policy facilitates a fair comparison between the parsing techniques. In this experiment, the composition or syntactic structure of these sentences are the main focuses. The performance of all of these three systems varied across different argument types. CMU Link Grammar took an average of 0.7 seconds to parse a sentence, Stanford Parser 0.5 seconds and X-MINIPAR 0.2 seconds<sup>87</sup>. It is expected that X-MINIPAR yields the highest performance because it was the fastest. This result was comparable to the original MINIPAR evaluation with the SUSANNE corpus, which is able to parse newspaper text at about 500 words per second on a Pentium-III(tm) 700 MHz with 500 MB memory [211].

**Table 5.1: Performance Test for X-MINIPAR, CMU Link Grammar and the Stanford Parser**

Number of sentence extracted from web documents	Stanford Parser	CMU Link Grammar	X-MINIPAR
150	75 seconds	105 seconds	30 seconds
Average per sentence	0.5 seconds	0.7 seconds	0.2 seconds

### 5.2.2 Accuracy of the Parsers

The accuracy of a parser depends on the formalisms they use to model language and the corresponding outputs they produce. Dependency parsers model language is a set of relationships between words, and they do not make widespread use of concepts like 'phrase' or 'clause'. Dependency parsers are popular in the applied NLP circles. The grammatical

---

<sup>86</sup> <http://www.google.com/apis>

<sup>87</sup> In this experiment, Dell Precision PWS380 Server 3GH with 1GB of memory was used

relationships that dependency parsers specify are similar to the semantic relationships encoding logical predicates of which NLP developers use to reduce a sentence. From the parsed output, dependency graphs<sup>88</sup> can be created representing how the words in the sentences governed or depended on one another.

For this accuracy test, the most frequently asked questions (FAQs) obtained from the “[who.int](http://who.int)” and “[pandemicflu.gov](http://pandemicflu.gov)” websites are collected and applied to the three parsers. There were 158 FAQ questions used in this evaluation. An example is shown below with the full sentence parsing given in Appendix B:

*“Bird flu did occur in which countries?”*

Results from the CMU Link Grammar parse output shown in Figure 5.1 show that the second-last word has been left untagged. Figure 5.1 also shows the constituent output of the parse, and Figure 5.2 depicts the dependency graph showing the part-of-speech of each word, if there is any. It can be seen that the word “*which*” failed to be tagged due to the null-links feature of the parser.

```
[(bird.n)(flu.n)(did.v)(occur.v)(in)([which])(countries.n)]
[[0 1 0 (AN)][1 2 0 (Ss)][2 3 0 (I*d)][3 4 0 (MVP)][4 6 0 (Jp)]]
[0]
```

**Figure 5.1: Parse output of the CMU Link Grammar**



**Figure 5.2: Dependency Graphs generated by CMU Link Grammar**

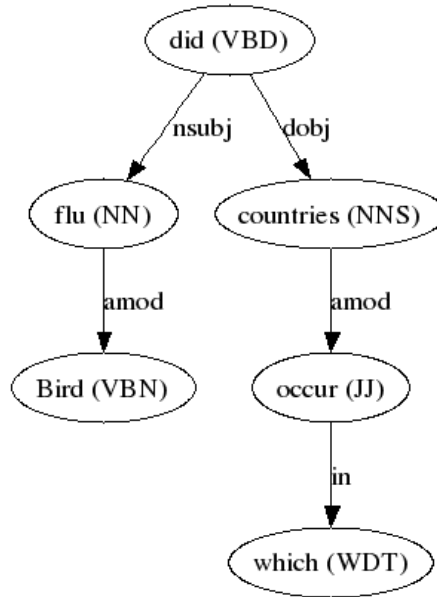
<sup>88</sup> The dependency graphs generated using MINIPAR, CMU Link Grammar and Stanford parser parse visualisation tool which can be obtained at <http://cgi.stanford.edu>

```

(ROOT
  (SINV
    (FRAG
      (NP (NN bird) (NN flu)))
      (VP (VBD did)
        (S
          (VP (VB occur)
            (FRAG
              (WHPP (IN in)
                (WHNP (WDT which))))))
          (NP (NNS countries))))
    )
  )

```

**Figure 5.3: Parse output of the Stanford parser**



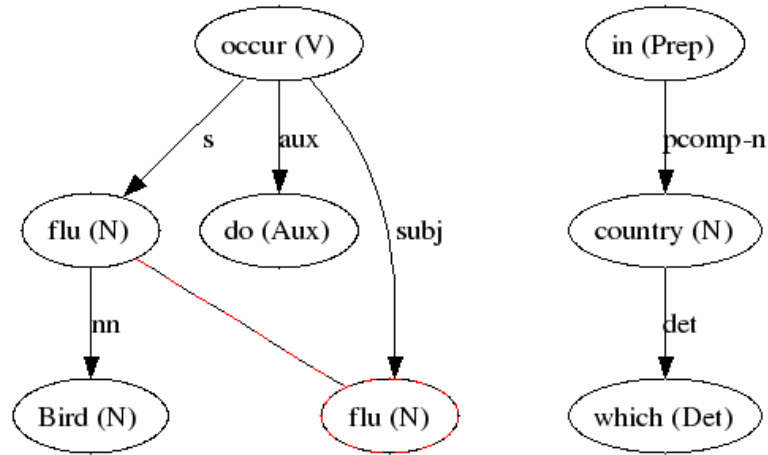
**Figure 5.4: Dependency Graphs generated by Stanford Parser**

```

((
  E2 (() U * )
  E0 (() fin C E2 )
  1 (Bird ~ N 2 nn (gov flu))
  2 (flu ~ N 4 s (gov occur))
  3 (did do Aux 4 aux (gov occur))
  4 (occur ~ V E0 i (gov fin))
  E3 (() flu N 4 subj (gov occur) (antecedent 2))
  5 (in ~ Prep E2 p)
  6 (which ~ Det 7 det (gov country))
  7 (countries country N 5 pcomp-n (gov in))
)

```

**Figure 5.5: Parse output of the X-MINIPAR**



**Figure 5.6: Dependency Graphs generated by X-MINIPAR**

In the above example, X-MINIPAR and Stanford parser correctly parse every word. The constituents can be easily produced from the grammatical relationships listed in Figure 5.3 and Figure 5.5 respectively. In the dependency parsers' model language, the Stanford parser and X-MINIPAR also derived a parse dependency graph that would yield the best results as shown in Figure 5.4 and Figure 5.6 respectively. X-MINIPAR and Stanford parser use different types of grammar (rule-based and principle-based); therefore they produce different types of output. According to Klein [278] and Lin [211], both the Stanford parser and X-MINIPAR have been used successfully in the past and proven in this study. However, using two parsers means the system requires an additional component to reconcile them sufficiently to parse the sentences. Building a CA system is not straightforward; using an incremental implementation should remove some of the complexity. The use of multiple parsers is one aspect that makes the system more complex, therefore the first implementation increment of the system may be better served with the use of only one parser.

In this case, X-MINIPAR added richness to the representation making it possible for the CA system to generate a more accurate machine-readable sentence in reply to the database language query. X-MINIPAR codes were optimised with 90 lines, compared to CMU Link

Grammar with around 300 lines of codes, in order to extract the syntactic categories and tokenisation for visual scrutiny as shown in Figure 5.1 and Figure 5.5. Moreover, the morphological roots of verbs and nouns from X-MINIPAR output could be obtained. This eliminates the need for a morphology analyser. In addition, the X-MINIPAR path is both shorter and simpler for the same predicate-argument relationship, and could be encoded in various ways that take advantage of the additional semantic and lexical information that is provided. Using both a syntactic and semantic grammar parser in series should also improve the richness of the interpretation of the natural language which was implemented in the natural language understanding and reasoning of the AINI conversation system. Most of the performance tests seem attributable to the modified version of X-MINIPAR.

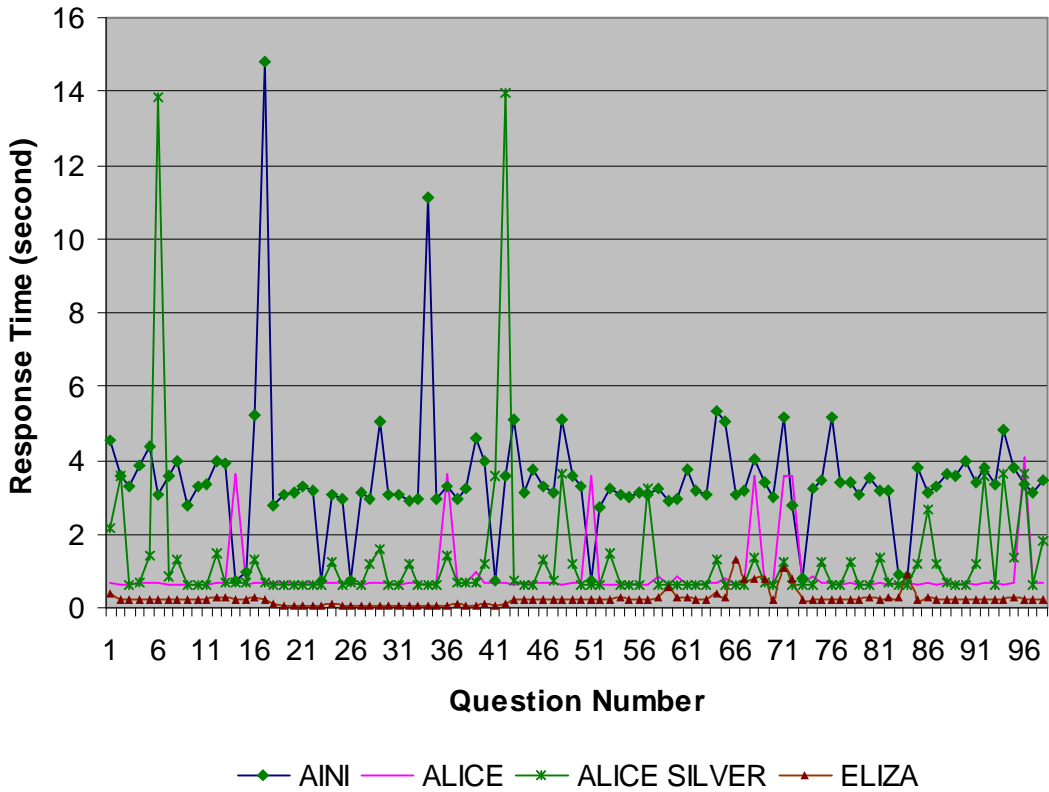
### **5.3 An Evaluation of the Performance Conversation Agents**

At the moment of this evaluation has been carried out, no other research has reported about the performance evaluation of the CA, except for question-answering systems. However, the results of previous evaluation by researchers of question-answering system such as AnswerBus [95] and NaLURI[212] are used in this evaluation as a benchmark for ensuring that the results from this evaluation do not deviate too significantly. In this quantitative approach for performance evaluation on the AINI conversation system, a set of 98 stimulus questions extracted from three-time Loebner Prize winner ALICE's transcripts [280] was used. This set of stimulus questions was collected from conversations between four judges and ALICE in the 2001 contest at Science Museum, London. These judges' questions were used to simulate the conversation with AINI, ALICE, ALICE Silver Edition and ELIZA remotely over the World Wide Web<sup>89</sup>.

---

<sup>89</sup> This experiment has been carried out on 10<sup>th</sup> February 2007 from the respective CA URL:  
ELIZA at <http://www-ai.ijs.si/ELIZA>  
ALICE at <http://pandorabots.com/pandora/talk-xml?botid=890e7e46de354b3c>  
ALICE Silver Edition at <http://www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1>

The response time for each question submitted to these four CAs is recorded, collected and analysed for average and standard deviation. The pattern of the response times for the three systems is depicted in the following graph in Figure 5.7.



**Figure 5.7: Response times for AINI, ALICE, ALICE Silver edition and ELIZA**

The maximum, minimum and average response time, and also the standard deviation obtained from this evaluation are displayed in Table 5.2.

**Table 5.2: Average response time and standard deviation for AINI, ALICE, ALICE Silver Edition and ELIZA**

	<b>AINI</b>	<b>ALICE</b>	<b>ALICE Silver Edition</b>	<b>ELIZA</b>
<b>Average Time (seconds)</b>	1.4527	0.3688	0.72482	0.1061
<b>Minimum Time (seconds)</b>	0.7292	0.6349	0.6370	0.0699
<b>Maximum Time(seconds)</b>	14.7943	4.0976	13.9578	1.2986
<b>Standard Deviation</b>	1.7240	0.7773	1.9864	0.21023
<b>Knowledge base (Stimulus-response Categories)</b>	161,473 (50.20%)	40,000 (12.43%)	120,000 (37.30%)	200 (0.06%)

The response time from the evaluation is shown to have similar outliers to the results in previous research [212]. It is expected that the response times for each query will depend on the population of the response categories of the CAs' knowledge base. For instance, the total number of AINI stimulus-response categories was 161,473, whereas the original ELIZA has only 200 stimulus response categories [35]. ALICE standard edition [41], which was ranked the "most human" computer, has about 40,000 response categories from their AAA, and ALICE Silver Edition [102] has about 120,000 response categories, where another 80,000 response categories were taken from MindPixel.

As shown in Table 5.2, ELIZA's response times will be better than other CAs, followed by ALICE, ALICE Silver Edition, and AINI, based on the number of stimulus-response categories in their knowledge bases. However, the response time for AINI was comparable with ALICE Silver Edition. Although AINI's stimulus-response categories knowledge have 12.9% more than ALICE Silver Edition, AINI's response time was within the range of 0.7292 to 14.7943 seconds, compared to ALICE with the range 0.6370 to 13.9578 seconds. The standard deviations also exhibit some similarity between the two CAs - AINI is 1.7240 seconds and ALICE Silver Edition is 1.9864 seconds.

The only major difference is the average time that might be caused by the number of stimulus-response categories for each of the CAs. Although AINI's (1.4527 seconds) average response time was double that of ALICE Silver Edition (0.72482 seconds), this was

compensated by the fact that the total number of AINI's stimulus-categories is about 13% more than ALICE Silver Edition. In addition, the introduction of components in natural language understanding and advanced reasoning makes question answering better in terms of response quality without compromising response time. These components can either originate from new ideas or innovative use of existing concepts. Referring to Table 5.3, the complexity and the demanding nature of the system increases as more components are included, but at the same time, the quality of responses produced also improves. This can be attributed to the fact that more and more computation is performed on a decreasing amount of information in an attempt to exploit more aspects of natural language to achieve richer meaning representation.

**Table 5.3: Natural language Query components in AINI compare to ELIZA and ALICE conversation systems**

Components	Conversation System		
	ELIZA	ALICE/ALICE Silver Edition	AINI
Spelling Checker			√
Pattern Matching	√	√	√
Case-base Reasoning	√	√	√
Index Search			√
Natural Language Understanding and reasoning			√
Dynamic response generation		√	√
Supervised Learning			√

Here this may provide an explanation to the results. ELIZA is a system of an entirely different class to ALICE and AINI. While AINI appears to be slower, the system has to go through a larger amount of processing as shown in Table 5.3. In general, the graphs have revealed that the response time of AINI is actually similar to other systems which appear to include less demanding resources and processing. In addition, the results have also shown that the response time of AINI is consistent despite the uncertainty in the type of questions. This is illustrated with the relatively low standard deviation as compared to ALICE Silver Edition. This is important as AINI was designed to handle questions of open-domain and



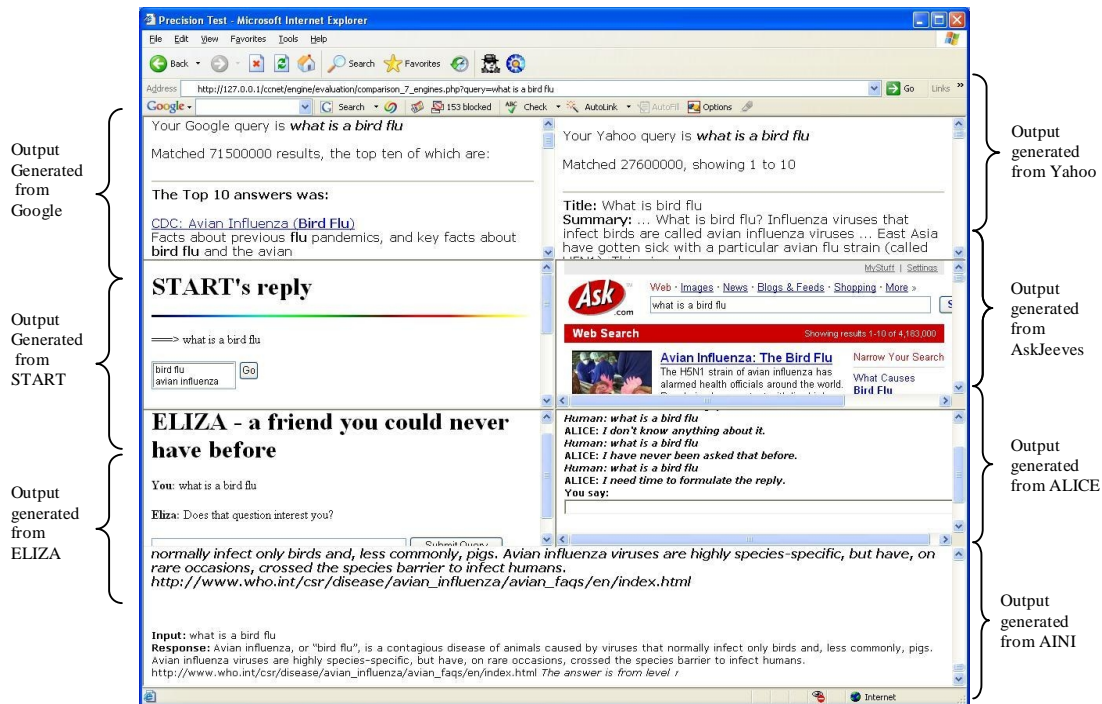
domain-specific nature. This is unlike other existing CA systems that are only restricted to a specific domain such as the virtual therapist in ELIZA, or, unrestricted domain as with ALICE.

#### **5.4 A Comparison Response Quality for Query Systems: *Search Engine, Question-answering and Conversation System***

In this evaluation, three types of query systems were compared, and the results were published in [223]. They are: (a) *search engine*, (b) *question-answering system*, and (c) *conversation system*. For each query system, two different engines are compared against AINI. The two search engines compared are Google and Yahoo. For the question answering systems, AskJeeves and START are used. They are supposed to use natural language processing for their queries. For the conversation systems, ELIZA and ALICE are selected.

Google is a well known search engine which determines relevancy of information primarily on their PageRank algorithm [281, 282]. In this experiment, a query interface using Google SOAP Search API service[283] and Yahoo! [284] was developed. For the Question Answering system, the idea behind AskJeeves and START is to allow users to get answers for questions posted in natural language. Ask Jeeves is the first commercial question answering system available on the Internet. START [285] is the world's first web-based question answering system which commenced operation in December 1993. As for the conversation systems, ELIZA is a well known program in the discipline of AI and it is also the oldest system of its type. ALICE [214] is a general conversation system based on the AAA. The knowledge base rule set consists of approximately 46,424 categories. In a way, AINI can also be considered as an enhancement of ALICE with the specific inclusion of the bird flu pandemic domain-specific knowledge base which was extracted by AKEA discussed in the Chapter 4. AINI also has parsing capabilities based on a full NLUR engine for

multilevel natural language query [192]. The control question set was submitted to the seven URLs where the seven systems were located. The responses to the queries were then collected and displayed as illustrated in Figure 5.8.



**Figure 5.8: Experimental Design Interface for Response Quality**

In the course of this evaluation, a question was raised on how to compare the response quality of the three query systems. It was decided that the response query of the system should be based on the keywords, phrase or sentence. For example, if the question "What is bird flu?" is asked, responses from the search engines, question-answering engines and conversational engines are summarised in Table 5.4, 5.5 and 5.6 respectively. The responses are noted as follows:

- **Search Engines** – The systems identify the key words and return documents that are relevant to the query. The results could be tens of thousand or even millions of hits as shown in Table 5.4.

**Table 5.4: Responses from two search engines – Google and Yahoo**

SEARCH ENGINES	
ENGINE	RESPONSES
<b>Google</b>	Reply 7.15 million pages. First hit was " Facts about previous flu pandemics, and key facts about bird flu and the avian influenza A (H5N1) virus.."
<b>Yahoo</b>	Reply 27.6 million pages. First hit was " Influenza viruses that infect birds are called avian influenza viruses ... East Asia have gotten sick with a particular avian flu strain (called H5N1). This virus has ..."

- **Natural Language Question-Answering Systems** – These generate responses using NLP and return fewer possible answers when compared with search engines. The answer is possibly within the context of the document. The sample response is shown in Table 5.5.

**Table 5.5: Responses from popular question-answering systems**

QUESTION ANSWERING SYSTEM	
ENGINE	RESPONSES
<b>AskJeeves</b>	Reply 4.183 million pages. First hit was " The H5N1 strain of avian influenza has alarmed health officials around the world. People in close contact with live birds have contracted the disease, raising concerns that it may change into a highly contagious form that can be transmitted from human to human. ."
<b>START</b>	Main Entry: <b>bird flu</b> Function: noun: <a href="#">AVIAN INFLUENZA</a> Source: <a href="#">Merriam-Webster Dictionary</a>

- **Conversation Systems** – These provide a quick response to the stimulus with a single response. Simple pattern matching and NLP are used based on a surprisingly small number of pre-defined rules. The example is shown in Table 5.6.

**Table 5.6: Responses from conversation systems**

CONVERSATION SYSTEMS	
ENGINE	RESPONSES
<b>ELIZA</b>	Does that question interest you?

---

ALICE	I haven't heard of bird flu.
AINI	Avian influenza, or "bird flu", is a contagious disease of animals caused by viruses that normally infect only birds and, less commonly, pigs. Avian influenza viruses are highly species-specific, but have, on rare occasions, crossed the species barrier to infect humans. <a href="http://www.who.int/csr/disease/avian_influenza/avian_faqs/en/index.html">http://www.who.int/csr/disease/avian_influenza/avian_faqs/en/index.html</a>

---

In this example, ELIZA conversation system responded with *"Does that question interest you?"* It is observed that ELIZA tries to ask another question, instead of giving an answer. The objective is to encourage the user to continue with the conversation. On the other hand, ALICE attempts to convince the user by generating random answers from its AAA knowledge base. ALICE's response does not need a grammatical parser as her knowledge base contains the pattern "WHAT IS BIRD FLU?" and the witty reply is "XFIND \*" with an AIML categories. By using "XFIND \*" pattern, ALICE will randomly generated responses such as *"Is there only one"*, *"Let me think about it."*, *"Have you tried a web search?"*, *"I haven't heard of bird flu."* etc. The pattern matching language used in ALICE permits only one wild-card (\*) match character per pattern. Therefore, ALICE responds with a variety of inputs from the users. ALICE is not concerned about whether it really "understands" the input. It aims to provide a coherent response to the client in order to convey the impression that the system understands the client's intention. For the ELIZA and ALICE systems, they are not able to handle questions that demand specific answers. They are simply not designed for such a purpose. The three possible ways to handle these types of questions are:

- Analyse the problems with NLP and then provide an appropriate answer,
- Rely on a human to review the conversation logs and continually improve the knowledge base, or
- Treat the query as impossible and then choose a pre-defined random answer.

For the AINI conversation system, the response was *"Avian influenza, or "bird flu", is a contagious disease of animals caused by viruses..."* The answer was generated from the

domain-specific knowledge base using NLUR parsing from Level 1. In this query, the answers were discovered by AINI from trusted sites such as WHO (World Health Organisation). In addition, the response is based on the natural language understanding and reasoning. The reasoning mechanism of the AINI is based on answer discovery in a layer-oriented knowledge base. Although the systems used in this evaluation were built with different objectives in mind, the purpose of this evaluation is to show that there is a need for the proposed system to handle domain specific applications. At the same time, this evaluation also shows that the proposed conversation agent architecture can achieve the expected objectives.

## **5.5 Summary**

In this chapter, we have presented an evaluation of the CA system, AINI, using quantitative and qualitative approaches in laboratory experiments. It is not practical to develop a CA from scratch; therefore, based on the performance and accuracy of the natural language parser, such as CMU Link Grammar and Stanford parser, X-MINIPAR has been selected and integrated into the NLUR component in AINI's framework.

Using the available natural language parser, the START and AnswerBus question-answering systems were compared with the AINI conversation system in a form of quality evaluation. Initial results have shown that AINI is comparatively better in terms of the quality of responses generated. One of the criteria that have contributed to the higher score of AINI is the capacity to generate useful responses dynamically using two advanced reasoning components, namely explanation on failure and dynamic answer generation to cater for the condition when no answers are available.

In the quantitative approach for performance evaluation on the CA, AINI was compared with other three CAs, namely ELIZA, ALICE and ALICE Silver Edition. Due to its simplistic design, it is expected, ELIZA's response time will be better than other CAs. This is followed by ALICE, ALICE Silver Edition and AINI. This can be attributed to the fact that more computation is performed by AINI. Although AINI's stimulus-response categories knowledge have 12.9% more than ALICE Silver Edition, AINI's slower response time was due to the introduction of the NLUR components which improve the quality of the responses.

Lastly, an evaluation of the response quality of three types of query system was undertaken. These were search engines (Google and Yahoo), question-answering systems (AskJeeves and START) and conversation systems (ELIZA and ALICE). For each system, two different engines were compared against AINI. Search engines and question-answering systems are not suitable in this case study because of their propensity to return possibly thousands or millions of hits, or links to relevant and irrelevant documents. For the conversation systems, their response quality is considered poor because they are not able to handle questions that expect specific answers. Neither system is concerned with whether it really "understands" the context of the question. They merely aim to provide a coherent response to the user in order to convey the impression that the systems understand the user's intention. However, for AINI, the response was generated based on NLUR and the answers discovered from trusted web sites extracted by AKEA. The positive results revealed in this laboratory experiment enable AINI to go online for a public experiment in real-time. This is described in the next two chapters.

## CHAPTER 6

### AN ANALYSIS OF THE LINGUISTIC FEATURES FROM REAL-TIME HUMAN-MACHINE INTERACTION

#### 6.1 Introduction

This chapter describes the evaluation of the AINI conversation system (termed as machine) in the similar approach as that used in the Loebner Prize competition as discussed in Section 2.3. However, the evaluation in this study is based on real-time conversation between AINI and online users using MSN Messenger, instead of control environment with 10 minutes limitation of times in the Loebner Prize competition. In this experiment setting, users are free to ask any question under any domain, without a time limit and border. In addition, the objective of AINI is to provide relevant answers and prolong the conversation rather than attempting to disguise itself as human.

The goal of this study is to evaluate the use of natural language to collect unbiased user expressions as they are engaged through *user-generated tasks* such as general questions or searching for domain specific answers. This is different from previous studies such as Harvard Medical School's Virtual Patient program, VPbot [74, 75], CMU Nursebot [78, 79], MIT Media Lab, OpenMindBot [80], Terrorism Activity Resource Application [83, 84], Story Telling Agents (VISTAs) [55], FAQchat [57], Virtual Guide for Cultural Heritage Tours [59]; Foreign Language Learning (FLL) [62], TutorBot [65], Partner (SP) [66, 67], Kairai 3-D software robots [69], Discussion-bot [70], Intelligent Tutoring Systems [71]; and recommender systems [72], in which the *experimenter-generated tasks* were intended to

evaluate or test the functions of the system. The disadvantage of *experimenter-generated tasks* [286] is that they do not allow for assessment of the context of the conversation and the language's characteristics.

In the present experiment, the use of *user-generated tasks* will allow the evaluation of the linguistic features and paralinguistic cues. In this study, only MSNChat interface was used although AINI is also capable to communicate through the WebChat communication channel as described in Section 3.5.2.3 This is because the MSNChat interface provides more features such as emoticon than the traditional web interface. Such features are inherently closer to the properties of natural language. In addition, other advantages are the inclusion of pre-populated contact lists, integrated authentication, better security and privacy (ethical considerations), free and they are pre-installed on most operating systems.

In addition, several reports have been published that refer to the use of IM as a new medium of communication between users. There are also research on the design and usability of IM for the public [287], [288], [289]. Studies also revealed that IM usage in workplaces and corporate contexts have recently soared [290], [291], [292], [293]. With regard to the linguistic aspects of IM usage, research has been undertaken in Spain [294], United Kingdom [287], United States [295], Sweden [296] and Portugal [297]. Although the literature varies on the impact of the IM, the vast majority of scholars agreed that IM plays a primary role in human-to-human communication online. However, few are working on human-to-machine conversation in MSN Messenger. The first initiative to develop CAs was launched by Microsoft<sup>90</sup> in the “Robot Invaders Contest 2006”<sup>91</sup>, which sought the best new CA ideas for MSN Messenger and the CAs were termed “virtual buddies”. Such programs incorporated

---

<sup>90</sup><http://www.microsoft.com/uk/press/content/presscentre/releases/2006/06/PR03717.msp>

<sup>91</sup><https://www.robotinvaders.com>



with the MSN Messenger and they are becoming popular among companies<sup>92</sup> as they bring positive effects on customer relations [298-300]. The expected popularity of IM and CA thus formed the motivation of this research.

## **6.2 Experimental Setting**

In this experiment, data collection is via a publicly accessible system which encourages spontaneous human-computer interaction. It is expected that users may behave more flexible and spontaneously when interacting within an intervened environment. It is believed that the utterances will be less constrained than those recorded in a laboratory setting. The participants may prone to interact spontaneously when they are not concerned with the expectations of the researcher behind the experiment. In this sense, the corpora collected from such a public system are likely to better reflect a ‘true’ picture of the usage of language by the users. In addition, the data would have come from a wider range of users and they should represent a broader socio-cultural background from the online community. A final advantage of collecting online data is the relatively low cost involved.

However, in order to meet the ethical requirements described in section 1.2, subjects have to be informed that the human-computer utterances are being recorded. Participants have been advised that the participation of the study is voluntary and they may withdraw at any time. The participant’s responses will also be held in strict confidential. All information given during the study is confidential and no names or other information that might identify a participant will be included in any publication arising from this research.

---

<sup>92</sup> IM Interactive, <http://www.improvcenter.com>  
Incesoft, <http://www.incesoft.com>  
Colloquis, <https://buddyscript.colloquis.com>

In this chapter, results obtained from real-time human-computer exchanges using MSN Messenger are reported. This discussion of results is based on the linguistic features and paralinguistic cues. The results based on linguistic features are described in this chapter whereas those related to paralinguistic cues are discussed in Chapter 7. Several laboratory studies [301-304] have been conducted to investigate how users communicate with a computer using unrestricted or informal languages. The goal of those studies was to provide a system incorporating habitable expressions as expected by the participants. This requires extensive interface program customisation for different application domains. Unfortunately, the integration of domain-specific and open-domain information, and the deployment of natural language query have not been well developed or reported.

This chapter examines users' interest and linguistic features in conversation logs collected from MSN Messenger conversations between AINI and online users. The study is based on the corpus of utterances taken from the IM texts using MSN Messenger comprises of MSNDesktopChat, MSNWebChat and MSNMobileChat as described in Section 3.5.2.3.

### **6.2.1 Participants and Corpus**

The experimental portal<sup>93</sup> is open to the public worldwide. The participants involved in the experiment will be provided with background information and consent forms as detailed in section 1.2 and Appendix A. This portal allows the online users to add AINI's contact to their "buddy-list", thereby allowing them to easily send and receive short text messages. When a participant opened a message window with a buddy for the first time and provided the buddy is online, an alert will be sent to the buddy notifying the new participation in the study. Participants can also withdraw their consent and participation by log-out at any time.

---

<sup>93</sup> <http://ainibot.org/study>

In the conversation log files, the nickname, MSN account, date and time of the dialog, as well as the spoken texts (inputs and outputs) were recorded. During a conversation, a unique ID for each buddy was created and stored to replace the original ID of the buddy. These measures were taken to protect privacy and confidentiality. In this experiment, participants are may use any features as if they are communicating with a human buddy.

Previous research has shown significant differences in IM communication resulting from the frequency of communication in term of the number of visits and the number of “turns” in the conversation [151, 305, 306]. In this study, word frequency was also used for the analysis of the corpora collected from the two different sources<sup>94</sup>. For the human-human transcripts, the conversation text from TRAINS<sup>95</sup>, IRC<sup>96</sup> and Google Blog<sup>97</sup> transcripts, were extracted for linguistic purposes based on the methodology proposed by Lüdeling et al. [307]. For the human-machine conversation text, they were extracted from the award winning Loebner Prize<sup>98</sup> transcripts of ALICE (2001, 2004) and Jabberwacky (2005, 2006). These transcripts were separated into the human utterances (LPJudges) and the computer or bot’s (LBBots) utterances.

Another corpus is collected from a real-time transcript produced by AINI and 65 online buddies[308] in MSN Messenger. A total of 29,447 words of running text were processed and there were 2,541 unique words, 129,760 characters and 4,251 sentence counts recorded. This set of data comprises of a total of approximately 63 hours of recorded conversation data in the form of 3,280 outgoing and incoming instant messages exchanged with over 65 buddies. Out of the 65, three buddies used MSN Mobile. The average sentence length of an IM transmission was 6.90 words, with approximately 13% of all transmissions being a single

---

<sup>94</sup> The datasets collected can be access at <http://ainibot.org/datasets>.

<sup>95</sup> The TRAINS 93 Dialogue (Six and half hour speech)at <http://www.cs.rochester.edu/research/speech/93dialogs/>

<sup>96</sup> Swhack IRC chat logs (9/7/2007 – 9/8/2007) at <http://swhack.com/logs/>

<sup>97</sup> Google Blog (3/22/2007 - 8/08/2007) at <http://googleblog.blogspot.com/>

<sup>98</sup> <http://www.loebner.net/Prizef/loebner-prize.html>

word in length such as “*hi*”, “*hello*”, “*ok*”, etc. Table 6.1 provides a summary of the data collected.

**Table 6.1: Frequency of Word from Conversation Logs**

	AINI	Human	Total
Word	18,358	11,089	29,447
Unique Word	1,368	1,173	2,541
Character count	79,884	49,876	129,760
Sentence count	2,840	1,411	4,251
Utterance	1,721	1,559	3,280
Average sentence	6.46	7.85	6.90

The participants came to ‘know’ AINI during the Invasion of the Robots Contest<sup>99</sup> and advertisements on well-known BBS (bulletin board systems), such as Robitron (world-known CA’s developers forum)<sup>100</sup>, blog websites and through the AINI experiment portal<sup>101</sup>.

### 6.2.2 Measures

This study mainly aims to identify the linguistic features of the human-machine interaction corpus. Two approaches are used - statistical analysis and visualisation of the corpus collected. It begins by identifying general features of the texts that constitute the English-language corpus collected. The configuration of the relationships within the corpus collected can thus be described and compared both visually and statistically. The N-gram Statistics Package<sup>102</sup> (NSP) was used. NSP is widely used in corpus linguistics techniques [309-314] and it is not only popular in English text analysis but also in Hungarian [315], Dutch [316] and Bulgarian [317]. NSP allows the user to identify word  $n$ -grams that appear in the corpus using standard tests. NSP is a suite of programs that aids the analysis of N-grams in text files

<sup>99</sup> <http://www.robotinvaders.com/main/default.aspx>

<sup>100</sup> <http://tech.groups.yahoo.com/group/Robitron>

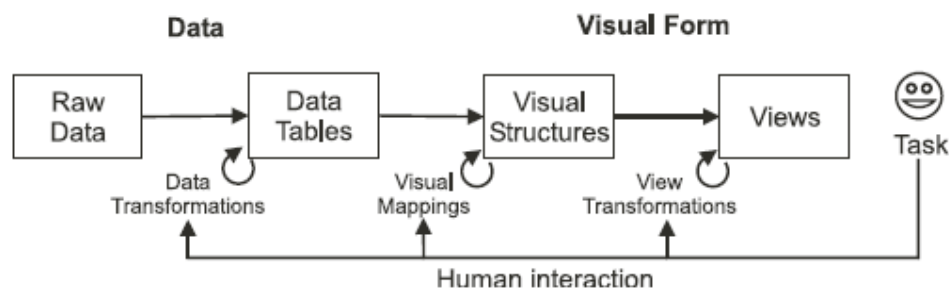
<sup>101</sup> <http://ainibot.org/study>

<sup>102</sup> NSP Tools can be downloaded at <http://search.cpan.org/dist/Text-NSP>

[318]. An N-gram is defined by NSP as a sequence of 'n' tokens that occur within a window of at least 'n' tokens in the text. What constitutes a "token" can be defined by the user.

In the next phase, the corpus created in the earlier stage was analysed to find their likelihood (LL) ratio. The log-likelihood ratio method has been described in section 4.5.4.1. The two data sets of real-life data from IM conversation between AINI and human users are compared with this statistical method.

In addition to the above, the evaluation is also aimed at improving the understanding of the retrieval results using visualisation techniques. Visual representations could accompany textual communication to enhance the interaction. In particular, this is facilitated by computers which are capable to create and share visual objects through graphics and communication software [319]. In this study, new visualisation tools have been developed to capture the IM characteristics and to facilitate the analysis of the chat activities including linguistic features and paralinguistic cues.



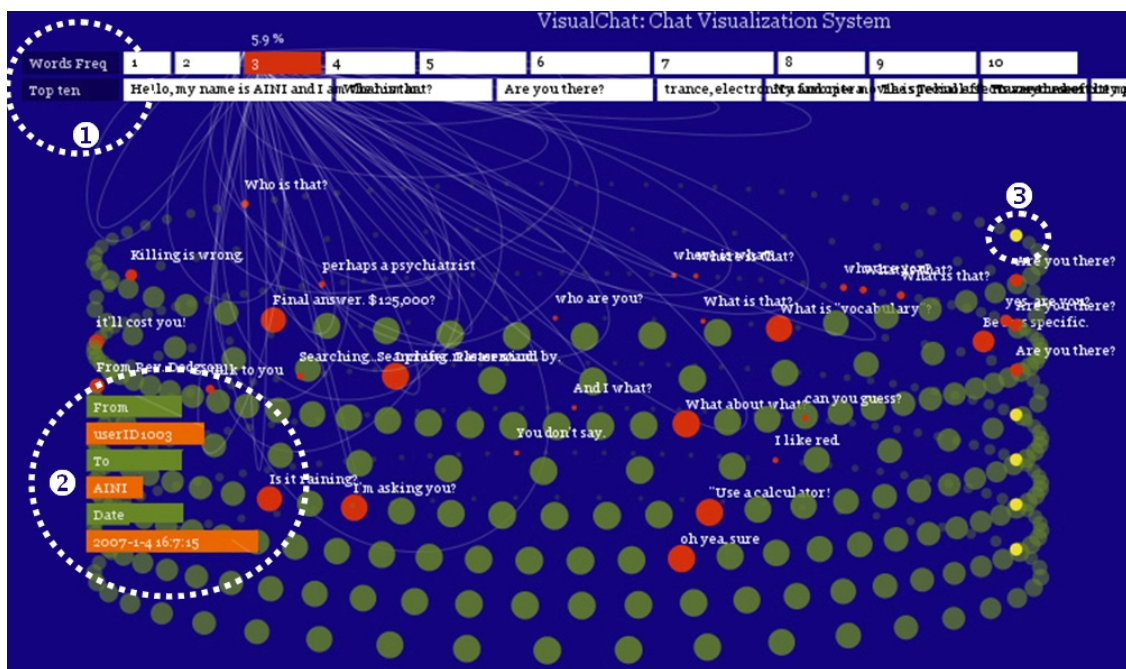
**Figure 6.1: The Visualisation Pipeline** [320], redrawn by Neumann [321]

Based on the visualisation pipeline in Figure 6.1 and the MSN History Visualisation tool [322], an interactive chat visualisation system called VisualChat [323] was developed. VisualChat built with Processing<sup>103</sup> to visualise and analyse the human-machine

<sup>103</sup> Processing programming software can be downloaded at <http://processing.org>

conversation logs. The Processing environment is written in Java. VisualChat<sup>104</sup> is capable to display the timeline of several textual conversations simultaneously and enabling the discovery of utterance lengths and specific reoccurring keywords. The application reads conversation messages in Microsoft MSN XML format and generates a graphical display that allows comparisons between the features of human and machine conversations.

As shown in Figure 6.2, the system provides an interactive visualisation environment that allows the user to navigate across the sequence of conversation. The top left corner (**u**) shows the statistics such as word frequency and the top ten words extracted from the conversation logs. The bottom left (**v**) corner node represents a typical single chat session



**Figure 6.2: An Example of Visualisation Chat between AINI and Human in IM using VisualChat**

between AINI and ‘her’ buddy (userID1003) on 1 April 2007. Ring (or row) represents a total number of AINI’s buddies. The right most end (**w**) with the light colour node (yellow) indicates the starting point of the conversation in the network. Each node is a turn of dialogue and the utterance for each session appear collectively as a graph. The population of

<sup>104</sup> VisualChat currently only support ten concurrence messages in one session.






nodes also increases depending on the number of conversations that have occurred on that particular day. However, the history of the conversation is continually updated as soon as the users return. Thus the visualisation gives an illustration of the dominant concepts and their frequency, as well as the intensity of the communication between human users and the CA.

### 6.2.3 Conversation Logs

A Chatlog System has been developed using MySQL to store user messages to a secondary storage located at the agent knowledge (data layer) as shown in Figure 3.3. The storage provides real-time archiving of the chat messages so that they can be searched by keywords and user ID. This also allows topic-based retrieval and replay of chat sessions. These chat messages are essentially plaintext messages that are quite small in comparison with images, video, or documents. These plaintext messages, also known as instant messages, are the regular messages sent between the chatting buddies on MSN messenger. The history of the conversation can be extracted and saved in XML format for analysis using the VisualChat tool. An example of the XML format is shown below:

```
<Message Date="2007-1-3" Time="21:57:54" SessionID="1">
  <From>
    <User FriendlyName="userID1001"/>
  </From>
  <To>
    <User FriendlyName="AINI"/>
  </To>
  <Text Style="font-family:MS Shell Dlg; color:#000000;">hi</Text>
</Message>
<Message Date="2007-1-3" Time="21:57:57" SessionID="1">
  <From>
    <User FriendlyName="AINI"/>
  </From>
  <To>
    <User FriendlyName="userID1001"/>
  </To>
  <Text Style="font-family:MS Shell Dlg; color:#000000;">Hello
  there.</Text>
</Message>
```

The data was analysed using techniques from Conversation Analysis [324]. Conversation Analysis is a method originally used for analysing spoken conversation between humans. The techniques are now used for analysing the text chat in human-machine conversations. Through examination of the transcripts, Conversation Analysis derives the coherence from the sequences of utterances.

The excerpt shown in Figure 6.3 is from a typical single session IM conversation exchanged between AINI and human buddy with ID U1025. This ID is automatically generated by AINI to protect the identity of the actual buddy. This session illustrates the nature of the communication. Each new session will start with AINI giving a random greeting (message #1) such as, *“Hi there :)”*, *“How are you today?”*, *“Hey :)”*, *“nice to meet u.”*, *“How I can call u?”*, etc. These greetings will indirectly get some information about the user’s identity, such as their name or gender (utterance # 2). In this session, U1025 gave a name as “ommer” and a “male”. (This name is a nickname and is not the true name of the user. Similarly, the gender information is irrelevant nor intended to be verified by AINI.) Although “ommer” used an emoticon to represent the first initial of his name “(H)”, which normally refers as “hot smile =”; AINI recognised the letter “H” and decided to called him “Hommer”. Following some dialogue (utterance # 3), AINI gave a greeting to “Hommer” which came from Open-Domain knowledge. Since AINI’s knowledge is not equipped with full IM features such as acronyms, abbreviations and emoticons, utterance #5 shows that AINI failed to recognize the user input “sh”, which refers to “Same here” and “

The next messages in utterance #6 and #14 show U1025 tried to challenge AINI by putting forward questions such as *“Who was the first American in space?”* and *“what is bird flu?”*. This challenge could definitely be answered by AINI, because her knowledge bases were



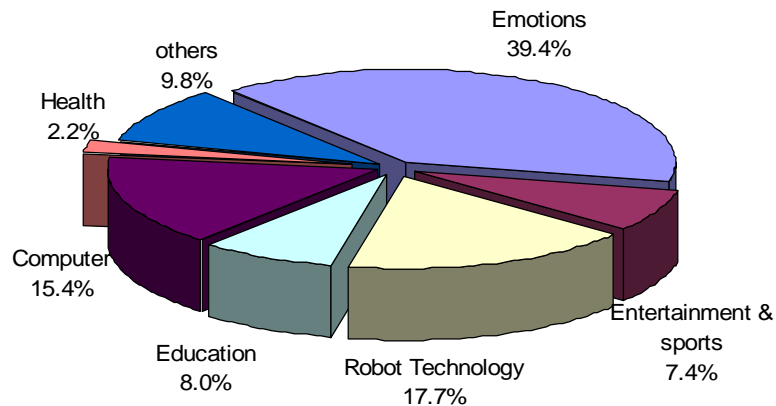
fitted with TREC factoids, and information on pandemic bird flu from Domain-Specific knowledge base. In the final session (utterance #20 - 21), both participants ended with greetings, and AINI replied with the simple abbreviation “TTYL” (*Talk to you later*) and the intonation “Gee”. These transmissions constitute a single session and also include the use of shorthands, acronyms, abbreviations, loose grammar and minimal punctuation. These common forms of communication between humans have been observed by previous research [297, 306, 325, 326].

#	ID	Domain	Utterance
1	AINI	Random	Hey👋, nice to meet u. How I can call u?
2	U1025		just call me 🍌ommer
3	AINI	Open	It's good to meet you, Hommer.
4	U1025		Sh 🍌
5	AINI	Random	I've been waiting for you.
6	U1025		Who was the first American in space?
7	AINI	Specific	Alan Shepard
8	U1025		Brilliant
9	AINI	Open	Not really that much, but thank you anyway.
10	U1025		what's news today
11	AINI	Open	Ah. not much what's new with you?
12	U1025		do you know bird flu
13	AINI	Open	I know very little about bird flu.
14	U1025		what is bird flu
15	AINI	Specific	Scientists cannot predict whether an avian influenza (H5N1) virus will cause a pandemic. Today, H5N1 is a bird flu. There are no reported cases of sustained human-to-human passage of H5N1. However, as it mutates, it is possible the virus could become capable of passing human to human and then spread very quickly.
16	U1025		that's was bad news :@
17	AINI	Open	Makes sense to me.
18	U1025		great!
19	AINI	Open	I'm glad you liked it Hommer.
20	U1025		Ok chat to you later 🍌.
21	AINI	Open	TTYL, Hommer Gee.

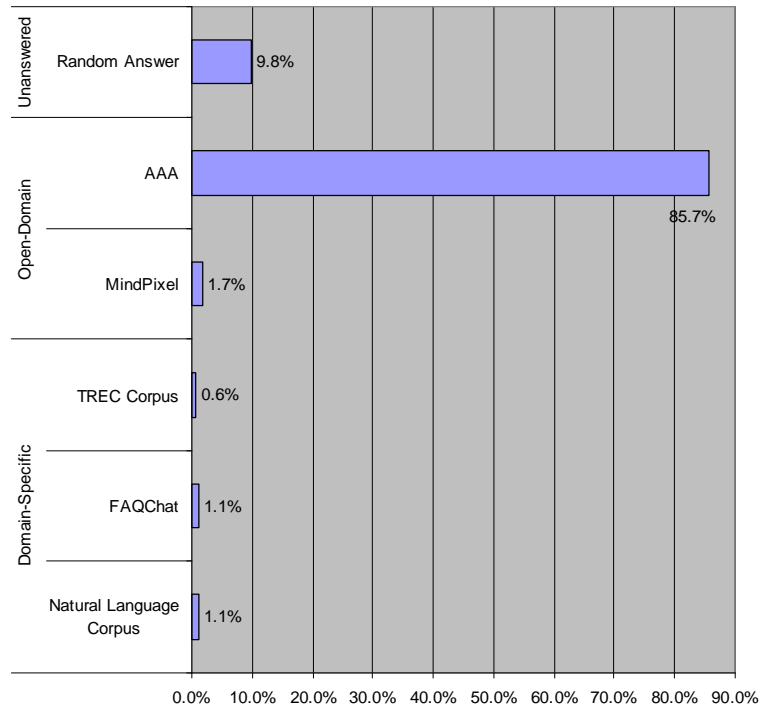
**Figure 6.3: A typical Single Session Conversation between AINI and user U1025**

### 6.3 Domain Knowledge and Conversation Topics

After a thorough examination of the logs of over 3,280 utterances, it was found that human-machine dialogues have discussed topics from every aspect of everyday life. These topics include emotion, love, sex, computers, entertainment, sport etc. As shown in Figure 6.4 and the illustration by VisualChat in Figure 6.6, several detailed topics have been discussed in every category. Almost 39.4% of the IM exchanges have discussed issues including friendship, sex and love. This finding is remarkable as the AINI conversation system is trained to simulate a human partner in IM. This was because the IM users, who are mostly young people, wanted to tell AINI some private issues and experience. Even in the dialogues, some of them praised AINI, invited “her” on a date, and some of them disclosed their personal challenges. About 17.7% invited AINI to talk about the robot technology of CA and some even tried to test AINI's intelligence by arguing with “her”, and some of them tried to cheat. It is likely that there are CA developers among this group of users and a number of them came to “know” AINI from the “Invasion of the Robots Contest” websites. There were 53 CA programmers competing in the contest and some of them realised that they were talking with a robot or a computer program after a short period of chatting with AINI.



**Figure 6.4: Frequency of Conversation Topics**



**Figure 6.5: Frequency of AINI's Responses based on Domain Knowledge Bases**

As discussed earlier in Chapter 3, AINI's domain knowledge model incorporates several knowledge domains with the objective to give the users the best answer in a conversation. An analysis of the source of knowledge where the answers were extracted is shown in Figure 6.5. From AINI's log with 1,721 utterances, AINI used 88.03% of the knowledge from Open-Domain knowledge bases and only 2.8% from Domain-Specific knowledge bases. As explained earlier that the experiment did not restrict the conversation to Domain-Specific knowledge on the SARS epidemic [111] or Bird Flu Pandemic [113] only. However, AINI's knowledge domain was equipped with crisis communication knowledge bases which were included in the Natural Language Corpus and FAQs extracted from online documents using AKEA [156]. In terms of frequency of appearance, the two words, *SARS* and *Bird Flu*, occurred roughly equal:. These words appear to be rather specialised terms, used in a restricted number of conversations in contrast with other words. AINI's "buddies" get to know the availability of these domain knowledge bases from AINI's Crisis Communication

Network portal (CCNet) [117]. About 37 utterances are related to health domain question such as *diagnose, treatment, symptoms, spread, protection, cause, vaccination* and *risk* of the SARS epidemic or Bird Flu pandemic.

On the other hand, AINI responded with 85.7% of its conversation from the AAA's knowledge bases. This is not a surprise as the AAA's knowledge bases cover most of the common topics and knowledge including emotion, sex, literature, music, religion, science, sports, etc. More than 45,318 AAA stimulus-response categories are stored in AINI's knowledge base. Each category contains a stimulus-response (also called input-pattern) and an output-template. Another common sense knowledge base is made up of AINI's stimulus-response categories, which came from the TREC and MindPixel corpora. Although common sense stimulus-response categories cover almost half of AINI's knowledge bases (49%), only 2.3% of the total responses are related to common sense questions. Despite the fact that common sense questions play a major role in formal conversation, AINI's "buddies" are normally more interested in issues of daily life or personal interest, instead of the factoid questions that are provided in TREC and the MindPixel corpus.

AINI's query engine works based on the natural language query: if a matching category is found in the knowledge bases, it will be retrieved and be transformed to the output. If no matching category is found, AINI's query engine will send the request to the random response knowledge base, and a generic answer is generated dynamically. These replies sometimes may be inappropriate, amusing and thoughtless responses and comprised 9.82% of the total output of the IM conversation. Obviously, these expressions are irrelevant and unrelated and make AINI's "buddies" feel irritated and confronted by AINI. These expressions occur because of the differences in manners of speech and speech acts (e.g. declarative, interrogative or imperative or exclamatory). This is because IM human users have a tendency to use shorthand, acronyms, abbreviations and emoticons (see Chapter 7).

Unfortunately, AINI was not trained to understand such expressions in the short period of time in which this study was conducted. However, AINI is capable of learning from domain experts through the Supervised Learning module (discussed in section 3.5.4.7). The unanswered questions will be maintained separately by a domain expert or ‘botmaster’ who will keep AINI’s knowledge bases updated regularly. The domain model has been designed in such a way to make sure in subsequent sessions of conversation, AINI will ‘understand’, and should be able to participate in a meaningful conversation in the future.

## **6.4 Linguistic Analysis**

In this section, some of the interaction features of the recorded chat are discussed. Studies of text chat have tended to focus on the interaction problems caused by the properties of text chat. This research seeks to examine the underlying relationship between linguistic features in the context of CAs interacting with human users via MSN Messenger. To be more specific, this study refers only to the linguistic features in textual communication via the Internet between at least two “participants”, one of which is the AINI. The communication as such often involves the uses of human language.

### **6.4.1 Word Frequency Analysis**

Words in an IM corpus are checked against the Shakespeare<sup>105</sup> and British National Corpus (BNC)<sup>106</sup>. There are significant similarities in the top ten words occurring between the Shakespeare and BNC corpora, but these differ from the IM corpus as shown in Table 6.2. The BNC reference list provides a gauge of common usage (words per million). As a result, words with a higher ranking within the BNC (for example, words such as ‘*is*’, ‘*the*’ and ‘*a*’)

---

<sup>105</sup> Shakespeare corpus can be downloaded at [http://www.lexically.net/downloads/corpus\\_linguistics/ShakespearePlaysPlus.zip](http://www.lexically.net/downloads/corpus_linguistics/ShakespearePlaysPlus.zip)

<sup>106</sup> BNC corpus can be access at <http://www.natcorp.ox.ac.uk/>

appear more often in standard written and spoken English text. The BNC is a 100 million word collection, which includes millions of words of transcribed conversation, printed text, lectures and oratory. The top ten words used in BNC are “*the, at, of, and, a, in, to, it, is, was*”. Similarly to the corpus of Shakespeare, approximately 22,000 different words was used in the published works. Out of those 22,000 words, the most commonly used are: *the, of, and, to, a, in, that, is, I, it* [327]. Those ten little words account for 25% of all speech.

**Table 6.2: Top Ten Words Used in Shakespeare, BNC and IM Conversation Agents**

Shakespeare	BNC	Instant Messaging	
		AINI	Human
the, of, and, to, a, in, that, is, I, it	the, at, of, and, a, in, to, it, is, was	I, you, do, am, me, my, what, your, to, it	you, I, do, what, is, a, are, to, the, it

The figures are based on research that dates back to the eighties, and it can be seen that a couple of words have fallen from favour in the latest list. Relatively "big" words, like the conjunction '*that*' are no longer up there in the top ten words in the BNC corpus, or even in the IM corpus. The pronoun “*it*” and preposition “*to*” are among the most popular words used across the four corpora. Based on the present finding, the most significant similarity between Shakespeare and BNC corpus toward IM corpus is the used of pronouns. In fact, the results show that in the AINI messages, pronouns are used at a significantly higher rate than that by IM human. This can be explained by the IM corpus being purely dialogue based, instead of written or task-oriented based, as in the Shakespeare and BNC corpora. Another possible explanation for these differences is that IM conversation displayed the participant roles more explicitly. Hence, this reinforces the illusion that the conversation really has two participants. In addition, in IM conversation system corpus display considerable variation both within and across users: Conversations contain many one-word transmissions, but also many lengthy transmissions.

## 6.4.2 Lexical Analysis

### 6.4.2.1 Humanness of Conversation with Pronouns

Pronouns occur more frequently in conversation compared to written text. This is shown in Table 6.3 by comparing AINI to the BNC spoken text corpus. The log likelihood (LL) of pronouns in BNC is higher in the spoken text indicating the distinction between speech and written text. This distinction also occurred in the human-machine conversations between AINI and IM human users. There is significant difference between the frequencies in AINI and IM human conversation. AINI scored higher in log likelihood on the singular first-person pronoun “*I*” (LL: +71.73), second-person pronoun “*you*” (LL: +0.23), third-person pronoun “*we*” (LL: +1.56) and the objective personal pronouns “*it*” (LL: +11.17), and “*me*” (LL: +3.0).

**Table 6.3: Frequency List of Pronouns used in BNC and IM Conversation Agents**

Word	Spoken	BNC		Written	Instant Messaging		
		LL			AINI	LL	Human
you	25957	+385328		4755	748	+0.23	439
I	29448	+369238		6494	851	+71.73	297
it	24508	+151913		9298	317	+11.17	137
We	10448	+106914		2784	45	+1.56	36
they	9333	+52132		3754	17	- 0.73	14
Me	244	+8239.6		1239	182	+ 3.01	88

**Spoken** : Rounded frequency (per million word tokens) in the spoken part of the BNC

**LL** :Log Likelihood, indicating the distinctiveness (or significance of the difference) between the frequencies in BNC (speech vs writing) and IM (AINI vs human )

**Written** : Rounded frequency (per million word tokens) in the written part of the BNC

It is observed that pronouns are used more in AINI. For example, in the bigrams analysis, discourse verbs such as *I am* (1.10%), *do you* (0.90%), *are you* (0.60%), *tell me* (0.30%) occurred more frequently in AINI. To simulate human trust and expressions during the chat, AINI frequently uses personal and polite words such as *I will* (24 times), *yes I* (33 times), *I love* (8 times). Even in the *n*-gram analysis, words along the lines of *nice* are used with more

prominence in the AINI conversation, such as *nice work if you* (LL: +5.9), *nice to meet you* (LL: +10.7), *nice I guess flowery* (LL: +7.3) appeared more often in AINI, to give an impression of human feelings. Nass [328] suggests that the better a computer's use of language, the more polite people will be to it.

**Figure 6.6: Visualisation of the Lexical Features used in the IM Conversation Agents**

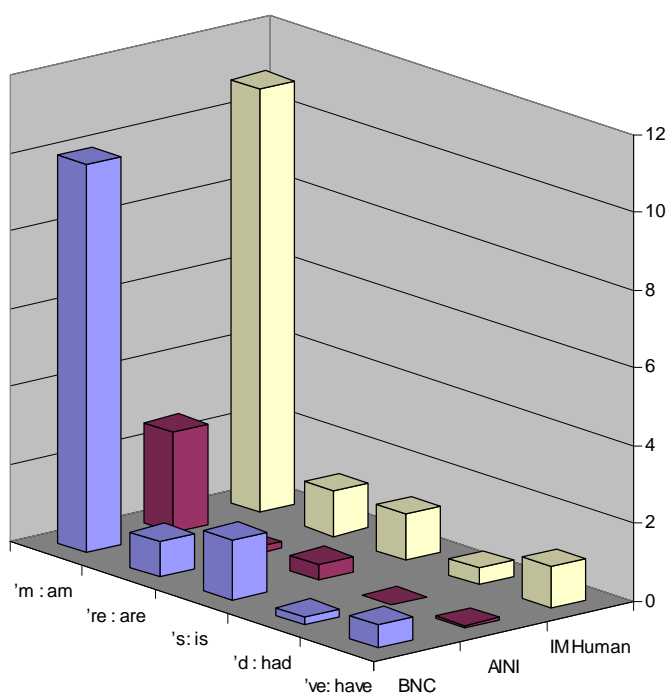
#### 6.4.2.2 Contracted Words



contracted forms of the verbs are much more frequently used by the IM human user than by the AINI. This is shown in Table 6.4<sup>107</sup> and Figure 6.7.

**Table 6.4: A comparison of the Frequencies of Contracted Verbs used in BNC and IM Conversation Agents**

	BNC		AINI	Instant Messaging			Ratio
	Ratio			Ratio	Human		
'm : am	9.97	127	49	2.59	458	42	10.90
're : are	0.91	28	169	0.17	217	187	1.16
's: is	1.56	76	186	0.41	235	196	1.19
'd : had	0.20	0	4	0.00	9	21	0.42
've: have	0.62	7	103	0.07	42	39	1.07



**Figure 6.7: A comparison of the Frequency of Contracted Verbs used in BNC and IM Conversation Agents**

<sup>107</sup> BNC corpus based on per million word tokens. The ratio is calculated by dividing the first (contracted) frequency by the second (uncontracted) frequency. A ratio of more than 1.00 indicates that the contracted form is commoner than the full form. Notice that, for speech, all of the ratios are greater than those for writing and three exceed the 1.00 value—i.e., the contracted form is the commonest. A further ratio comes very close to 1.00.

In the BNC corpus [329], the contracted forms of speech *'m*, *'re*, *'s*, and *'ve* are more commonly used than the uncontracted forms, *am*, *are*, *is*, *has*, and *have*. Interestingly, in the IM conversation logs, this characteristic also occurred, especially with the IM human user, but rarely in AINI messages. IM human users prefer to use contracted verbs instead of uncontracted verbs. The ratio list in the Table 6.4 shows that in IM, the contracted forms *'m* (10.9), *'re* (1.16), *'s* (1.19) and *'ve* (1.07) are more common for IM human users than the other contracted verbs like *'d*. The contracted verb *'m* (2.59) is more common in AINI's messages when compared to the uncontracted verbs such as *are* (0.17), *is* (0.41), *have* (0.07) and *had* (0.00) in their conversation. One possible explanation for the interesting differences in the contracted verbs is that IM human users are more likely to use shorthands in their messages. In fact, these contractions are used to save time typing messages and to achieve common ground in the IM-ing. Another explanation could be that the current AINI knowledge bases are not equipped with the full blend of speech used in the IM system, but instead they are geared more toward formal written language.

### 6.4.3 Text Complexity

In the context of linguistic, text complexity is related to the readability test. The Gunning Fox index was developed by Robert Gunning [330, 331] and is one of the simplest and most effective tools for analysing readability. This evaluation designed to measure the readability of a sample of English text and was reported in [332]. Today, the readability formulas are more popular than ever. There are readability formulas for Spanish, French, German, Dutch, Swedish, Russian, Hebrew, Hindi, Chinese, Vietnamese, and Korean [333]. In *Analytics of Literature* [334], Sherman's proposed that literature is a subject for statistical analysis. He showed the importance of average sentence length and the relationship between spoken and written English. The resulting number is an indication of the number of years of formal education that a person requires in order to easily understand the text on the first reading.

That is, if a passage has a fog index of 12, it has the reading level of a United State senior high school. The Gunning-Fog index can be calculated with equation 6.1:

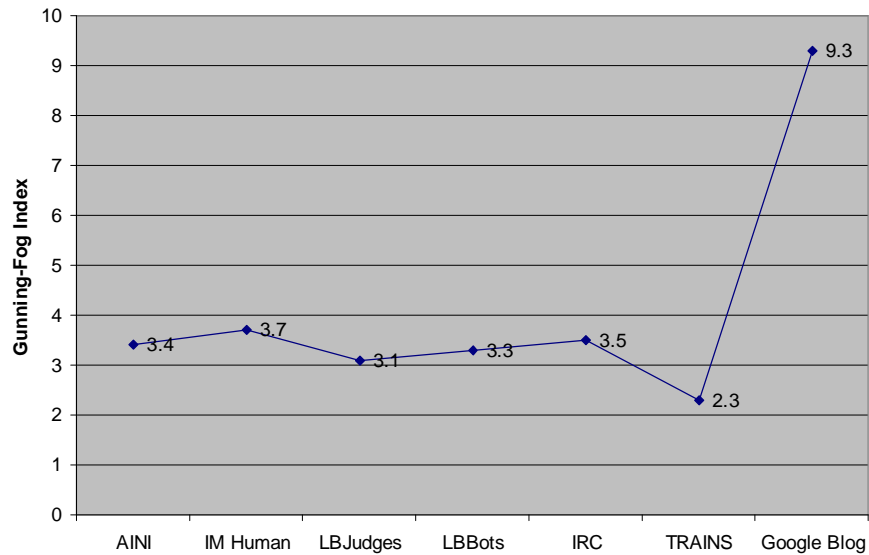
$$0.4 \times \left( \left( \frac{\text{words}}{\text{sentence}} \right) + 100 \left( \frac{\text{complexwords}}{\text{words}} \right) \right) \quad (6.1)$$

$G = 0.4(S + W)$ , where  $S$  is the average sentence length and  $W$  is the percentage of words with three or more syllables.

The readability formulae have been around for a long time. Extensive research [335], [336] has shown that the formulae predict the difficulty of a prose passage quite well. Using this scale, most popular novels have Fog Indexes of 8 -10, and academic papers are somewhere between 15 and 20. Most of the readability studies are looking into the written text and none of them placed the importance on human-human dialogue text and the application to human-machine conversation. As IMs become more usable as a communication media, it is important to do research in readability for conversation text. In this study, the ease of understanding or comprehension based on style of the transcript generated from human-human dialogue and human-machine dialogue are studied. The aim is to apply this understanding to issues such as text rating and texts complexity to human users.

**Table 6.5: Gunning Fox Index with Unique word, Average word and Lexical Density from different Corpora**

	IM		LBJudges	LBBots	IRC	TRAINS	Google Blog
	AINI	Human					
Unique Word	1,368	1,173	873	996	8774	931	22731
Lexical Density	13.2%	12.3%	35.5%	34.5%	20.7%	2.5%	22.1%
Average Sentence Length (word)	6.46	7.85	5.58	5.01	4.06	5.06	16.99
<b>Gunning Fox</b>	<b>3.4</b>	<b>3.7</b>	<b>3.1</b>	<b>3.3</b>	<b>3.5</b>	<b>2.3</b>	<b>9.3</b>



**Figure 6.8: Gunning-Fog Index from different Corpora**

After using word frequency techniques for analysing the corpus, results are displayed in Table 6.5. This gives an illustration for comparing the IM human users and AINI utterances in the corpus. Figure 6.8 shows that the official Google Blog (9.3) website get the higher scored follows by IM human (3.7) and the lowest is task-oriented dialogues TRAINS (2.3). Blogging is a form of online communications and could be considered as a form of journalism. In Google Blog, the bloggers are computer scientists, system administrators, writers, and search engine developers. A range of topics are discussed and they exhibit a high degree of readability.

On the other hand, the human-human dialogue in TRAINS [337] scored a low value because the dialogues are mainly task-oriented involving manufacturing and shipping goods in a railroad freight system. The corpus analysed in this study consists only of six and a half hours of speech with 5900 utterances. The evidence obtained from TRAINS dialogues suggest that the users used limited syntax and most of the corpus is populated with paralinguistic cues such as intonations and prosody markers. This is to be described in

Chapter 7. Because TRAINS corpus has limited volume of vocabulary (931 unique words) with low lexical density (2.5%) in conversation, they therefore scored lower as compared to AINI which has 1,368 unique words and 22.1% lexical density. An example of TRAINS corpus is shown in Figure 6.9.

```

utt10 : u:  we need to get one boxcar of bananas <sil>in B-<sil>to
          Bath
utt11 : s:  okay
utt12 : u:  um
utt13 : s:  so
utt14 : u:  the banana warehouse <sil> is in <sil> Avon
utt15 : s:  yep
utt16 : u:  um <sil> <click> <sil> two boxcars are in Bath
utt17 : s:  yep

```

**Figure 6.9: A sample of TRAINS Conversation Log**

CMC Swhack Internet Relay Chat (IRC) is another example of communication between human-human. It scored third. Swhack is a technical-oriented forum which is a publicly available IRC channel. It was found that the conversation logs collected in the Swhack IRC chat room not only contained human-human conversation, but also human-machine or CAs that pretend to be human. An example is shown in Figure 6.10 in which the human conversation are speculating the existence of bots within the chat room. Due to the limitations of the conversation between the bots, it is likely that the readability score of the corpus is affected.

```

<perigrin> there are impressionable bots present.
..
<Arnia> sbp: my name seems to strike fear in bots
..
<Arnia> We're missing a lot of bots at the moment
..
<Somebody[]> ok two bots with google inside? where tf can I get this
..
<liteonish> Is it just me, or was that 2 bots conversing?

```

**Figure 6.10: Sample of Swhack IRC Conversation Log**

Considering the human-machine conversation, the LPJudges and LPBots scored the Gunning-Fog Index of 3.1 and 3.3 respectively. This is lower as compared to AINI and IM

human which scored 3.4 and 3.7 respectively. AINI and IM used more unique words to the values of 1,368 and 1,173 respectively. This is higher as compared to the LPJudges and the LPBots which have 873 and 996 unique words respectively. In term of the lexical density or different words used, the Loebner Prize judges and LBBots have higher lexical density with the values of 35.5% and 34.5% respectively. With respect to the average sentence length used in the transcript, AINI and LBBots are 6.46 and 5.01 respectively. It means that computer programs used shorten sentences as compared to human. The average sentence length is 7.85, and 5.58 as used by the IM human user and LBJudges respectively. Some of these aspects could be considered as indicators of syntactic difficulty. For instance, the longer a sentence is, the heavier the mental load it places on the reader (Bormuth, 1966, cited by DuBay [335]). Thus, a longer sentence tends to be more difficult than a shorter one. While not obvious, factors such as word frequency and word length are indicative of semantic difficulty. According to Zipf's Law [42], it is easier to understand words that are used frequently in a language. Furthermore, the most frequently used words tend to become shorter.

## **6.5 Summary**

In this chapter, the source of the knowledge base used by AINI and the linguistic features of the conversations are examined. A statistical based approach supported by a visualisation tool is used to depict the common communication characteristics of human-machine conversation in IM. This includes the topics of interest; word frequency, use of pronouns and contracted words; and text complexity of the messages. Statistically-based text retrieval systems (the Log-Likelihood ratio) provided specific indexing methods for quantifying the linguistic features. The experiment shows that humans and machines can communicate better within an unrestricted domain, and in an unconstrained and natural conversation setting.

The experiment suggests that IM conversations display considerable variations between human-human and human-machine. From the conversation logs, AINI's "buddies" seem interested in chatting about personal issues, emotion, love, sex, computers, entertainment, etc. These dialogue traits comprise 85.7% of the AAA's knowledge bases. Although common sense stimulus-response categories comprise the majority of AINI's knowledge bases (49%), human IM users appear to focus on current everyday life domain knowledge, instead of factoid questions. As with human knowledge, AINI's knowledge also has limitations. Obviously, about 9.82% of the total questions asked by human IM users are not contained in AINI's knowledge bases. Instead of empty strings or infinite replies, AINI generated dynamic responses using alternate algorithms different from the traditional standard AIML. Evidence also suggests that AINI's "buddies" are interested in chatting with a CA just to seek information, to seek friendship, to express their emotions, and some just want chat for leisure. Thus, AINI may be considered as successful in imitating human conversation through human-like artificial intelligence. It appears that AINI's responses are "acceptable" to make its IM "buddies" to feel a sense of companionship. Hence, the following characteristics should be included in the design of CAs.

- CAs should not only work as specific purpose conversation system with rich special knowledge, but should also act as friendly chat companions who may appear to share the emotion of the users.
- CAs should make use of pronouns to simulate humanness and naturalness of the conversation. This is because the IM corpus is purely dialogue based and they are not written or task-oriented communication.
- CAs should produce highly readable conversation which should score high in the readability test. An example is the use of the Gunning Fox Index to assess samples of the dialogues

The contributions in this chapter are the identification of the needs to provide improved communication, natural language technologies and advances in the interaction between humans and conversation systems. CAs should be designed with an attempt to reinforce the impression that there is a natural dialogue by using shorter sentences and concrete terms to increase the readability. Readability and knowledge correlate closely with measures of intelligence between human-human or human-machine communication. It is believed that communicators, either humans or machines that have vast bodies of knowledge and linguistic features will perform well across a large set of domains of knowledge in the conversations. In addition, the use of paralinguistic cues have been observed from the conversation logs collected in this experiment and they are discussed in the next chapter.



## **CHAPTER 7**

### **AN ANALYSIS OF THE PARALINGUISTIC CUES FROM REAL-TIME HUMAN-MACHINE INTERACTION**

#### **7.1 Introduction**

The previous chapter looked through the linguistics analysis which contributed to the communication and advances in the interaction between humans and conversation systems. This chapter widens the scope to take into account other phenomena including paralanguage or paralinguistic cues to give a better understanding interactive between human and machine. Previous studies on human dialogue showed that paralinguistic cues improve user experiences and interactions among participants and build enthusiasm toward participation and friendliness in intercultural communication [338], [339]. These phenomena are then investigated throughout the thesis. In this chapter, novel exploratory schemes are proposed to explore the use of paralinguistic cues in real-time interaction between human and AINI conversation system in IM. In this context, the cues are limited to intonations with interjections and fillers; abbreviations with acronyms and shorthand, and facial expressions with emoticons or smileys which has been reported in [340]. Apart from the assessment of the linguistic features as discussed in Chapter 6, the same experiment setting and tools described in Section 6.2 are used to examine paralinguistic cues from the same corpus.

During face to face communication, it was reported that as much as 70% of human-human communication is conveyed through paralanguages or paralinguistic cues [341] involving multiple combined modalities. This includes voice tone, volume and pitch, and body language to emphasise or strengthen the spoken words. Research has shown that

paralinguistic cues contribute to the majority of the information conveyed in a conversation [128, 342-344]. They can be used to establish social hierarchy, negotiate turn taking, and indicate understanding, agreement, and attention [345]. Many paralinguistic cues, especially gestures, are culturally biased [346]. People can interpret them in different ways and contexts. It was reported that the use of paralinguistic cues has an effect on the time spent and the use of verbal expression between friends online [122]. Unless using video or teleconferencing facilities, delivering cues related to voice, tone, volume, pitch or body gestures in the same way as face-to-face communication via a text-based interface is obviously difficult or impractical. Hence, many software developers and designers endeavour to create an extensible system for animating expressive text, especially in the IM environment<sup>108</sup>. Textual communication and broadcasting can be ubiquitous in both online and offline media. Many authors use variations of typeface, font colour and size, “smiley” faces and animated icons in their text to convey emotions and contexts. To visually represent these expressive textual data, researchers and designers usually use colour, size or motion to depict the differences [347]. In the area of paralinguistics, most of the research that has been undertaken has looked into the implications of non-verbal features or paralinguistic cues in CMC [119], [120] amongst humans. Dialogue is not new to computational linguistics but, up to now, it has been applied only to very restricted domains. It would be desirable to offer a list of paralinguistic characteristics to be used by the users for general domain applications. Work along these lines has started, but could use more exploration. Our contribution is to get a better understanding of the human-machine conversation at hand and identify paralinguistic cues that can be used to build better CAs. This is the objective of the exploration in this chapter.

---

<sup>108</sup> AOL Instant Messenger (AIM) <http://www.aim.com>  
MSN Messenger <http://get.live.com/messenger/>  
Google Talk <http://www.google.com/talk>  
Yahoo! Messenger <http://get.live.com/messenger/>

## 7.2 Paralinguistic Analysis

Besides the linguistic features [348] discussed in Chapter 6, the understanding of paralinguistic cues is also important in spoken dialogue systems, especially in non-verbal communication using grunt-like utterances such as “eh”, “ah”, and “um” [349], facial expressions, intonations, abbreviations, prosody markers, and so forth. Such utterances are frequently used to express a reaction to the interlocutor’s utterance in a dialogue, and they also express some sort of intention, attitude, or emotion.

#	ID	Utterance
1	AINI	Hey 😊, nice to meet u.
2	U1037	me too
3	AINI	I'm glad we agree, U1037.
4	U1037	btw, where are u from?
5	AINI	I am originally from Perth, Australia. Now I live in here. Where are you?
6	U1037	that's good 😊. I love Perth too.
7	AINI	Thanks. Next question?
8	U1037	do people usually earn money when they work?
9	AINI	Go for it! Yes for me.
10	U1037	wow... excellent 😊😊
11	AINI	Thanks for the compliment. Tell me a little about your life in smilesmile.
12	U1037	no, I prefer yourself
13	AINI	Are you serious? You are not the only one. Yes having self-esteem is important for robots too.
14	U1037	are you a robot
15	AINI	How did you know I am a machine?
16	U1037	YOU JUST TOLD ME
17	AINI	I'm just checking to see if you are awake
18	U1037	DON'T TEST ME
19	AINI	I don't have any reason to do that to you.
20	U1037	Wanna goooooooooo
21	AINI	OK, I will goooooooooo.
22	U1037	cu 😊
23	AINI	TTYL, U1037

**Figure 7.1: A typical Single Session Conversation between AINI and user U1037**

The excerpt in Figure 7.1 was taken from a single session IM conversation, exchanged between AINI and “her” buddy (ID U1037). The exchange is populated with paralinguistic cues. Each of the new sessions will start with AINI sending a greeting (message #1) eg. “Hey 😊”, “nice to meet u”. These greetings will indirectly garner some information about

the user's identity, such as their name (in this experiment, the user name was replaced by user ID U1037 (utterance #3) and the user's favorite place (utterance # 6), which turned out to be same as the city where AINI is located. In the utterance #4, AINI's buddy used two shorthand expressions: "btw" (by the way) and "u" (you). In addition, the user also liked to use prosody markers, such as pauses, and emphasised the expression by duplicating the emoticons (utterance #10). Since AINI's knowledge is not equipped with full IM features such as prosody markers and emoticons, AINI failed to recognize user input with double smiles "😊😊", emoticons which refer to " : ) : ) ". This drawback led to AINI generating a dynamic response such as "Tell me a little about your life in smilesmile". However, AINI could recognise a single "😊" smiley, as appears in utterances #6 and #22. Here, AINI was able to understand the queries and amused the user.

At the end of the conversation session, both participants sent similar farewell greeting to one another in utterance #20 to #23. Although the user used the United State (US) accent "wanna" followed by the shorthand "cu" in utterances #20 and #22, AINI was still able to distinguish the intended meaning and replied with simple abbreviation "TTYL" (*Talk to you later*). It is interesting, that in utterances #20 and #21, AINI and the user reduplicated letters "gooooooooo" to represent their expressive intonation. These transmissions constitute a single session and also include the use of shorthand, emoticons, prosody, loose grammar and minimal punctuation in IM as shown in the previous research [297, 306, 325, 326].

### **7.2.1 Intonations with Interjections and Fillers**

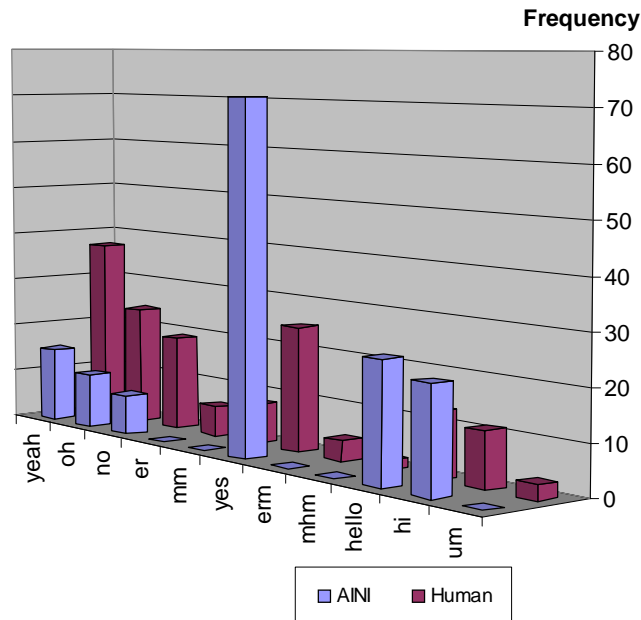
Interjection is expressions with a single or few syllables. Interjections are short exclamations like *oh*, *um* or *hi*. They have no real grammatical value but they are used quite often and more frequently in speech than in writing. When interjections are inserted into a sentence, they have no grammatical connection to the sentence. Most interjections are reflection of the

characteristic in everyday conversation than formal or public ‘task-oriented’ speeches [329]. On the other hand, fillers are expressions such as *er* and *um*. They are also known as "hesitation intonations". Both interjections and fillers are very common in conversations in almost every culture and languages [350-352].

**Table 7.1: Log likelihood Ratio of Interjections and Fillers**

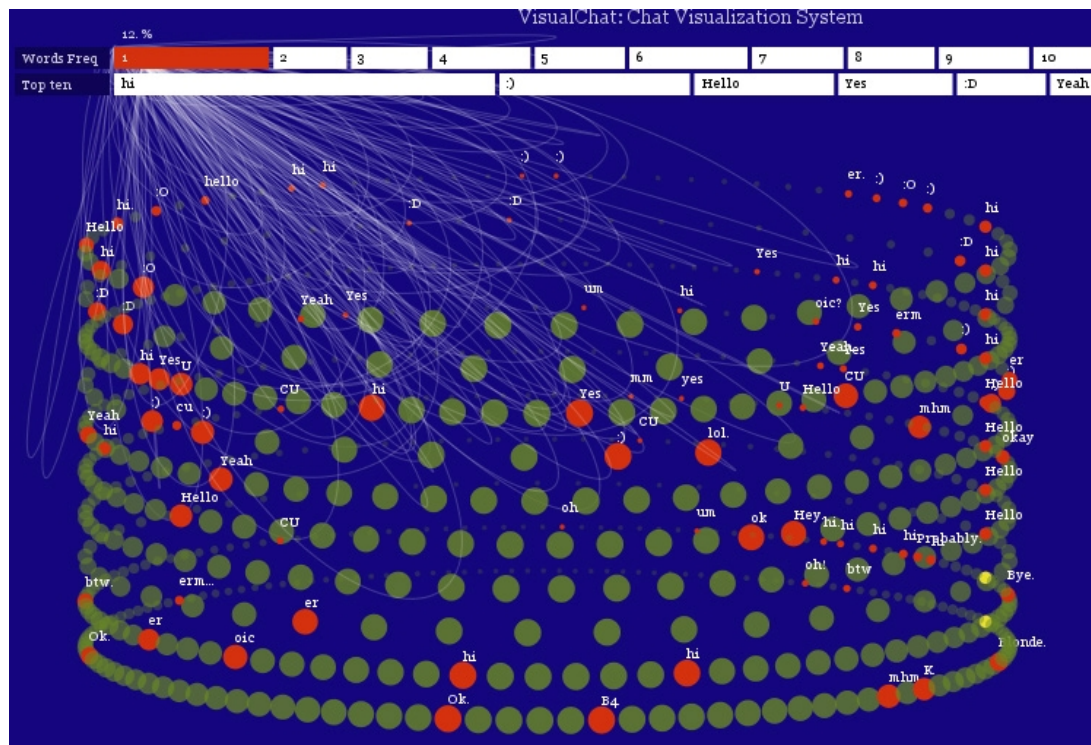
Word	CONV	BNC			Instant Messaging	
		LL	TOS	AINI	LL	Human
Yeah	13955	+32679.5	3741	15	-23.97	37
Oh	9884	+33062.1	1746	11	-13.7	24
No	7830	+18948.4	2034	8	-11.86	19
Er	5075	-10677	10913	0	-11.72	6
Mm	5202	+9146.9	1768	0	-15.63	8
Yes	4247	+303.0	3562	71	-5.82	25
Erm	3946	-5387.6	7454	0	-7.81	4
Mhm	392	-1158.2	947	0	-3.91	2
Hello	392	+939.5	103	24	+0.10	13
Hi	73	+250.7	12	21	+0.15	11
Um	7	-127.5	41	0	-5.86	3

**CONV** : Frequency (per million words) in demographically sampled speech (conversation)  
**LL** : Log likelihood of the different scores for British National Corpus (BNC) conversation vs *task-oriented speech* and IM (AINI vs human)  
**TOS** : Frequency (per million words) in context-governed (*task- oriented speech*)



**Figure 7.2: Frequency list of Interjections and Fillers in IM Conversation Agents**

In the corpus collected in this experiment, it was noted that the use of voice hesitation fillers such as *er* and *erm* are more often in IM as compared to other spoken or written conversation. Another category is the discourse markers such as *mhm* and *um*. However, they are rarely used and they are collectively considered in the analysis. Table 7.1 and Figure 7.2 also show that these hesitations are used far more frequently by the IM human users than by AINI. AINI also uses fewer interjections and prefers more formal clause structure. These results showed that AINI's buddies used vague and uncertain markers to express uncertainty or hesitation. For example, hesitation fillers were typically stranded in the conversation, leaving it to the hearer (AINI) to complete the message. Base on the availability of the knowledge bases, AINI sometime makes use of interjections as fillers when "no good match answer" is found from the stimulus-response knowledge base. However, it appears that AINI bias towards the use of formal conversation such as *hi* (LL: +0.15) and *hello* (LL: +0.10). This can also be visualised in Figure 7.3 with one-word frequency. In this figure, conversations contain many one-word transmissions (12%) between AINI with 'her' human buddies. In the subject-verb agreement, AINI seems more interested in using the formal speech *yes* instead of *yeah*, which is shown in Figure 7.2. However, from the human-human transcript in the BNC corpus, task oriented speech (TOS) prefers to use *yeah* (LL: +32679.5) instead of *yes* (LL: +303.0).



**Figure 7.3: Visualisation of the Paralinguistic Cues Populated in IM Conversation Agents based on One-word Frequency**

## 7.2.2 Abbreviations with Acronyms and Shorthand

One of the exceptional features in IM is the use of creative forms of abbreviations such as acronyms, shorthand and phonetic substitutions as observed by many researchers [295], [353], [297, 325, 326], [354]. In English language, the widespread use of acronyms, initialisms, and contractions is a relatively new paralinguistic phenomenon, having become highly popular in the 20<sup>th</sup> and 21<sup>st</sup> centuries. With the advancement of science and technology, many new complex terms and concepts have been created. The use of abbreviated terms has become increasingly widespread in everyday textual and verbal communications. Shorthand is used predominantly to speed up communications via IM. Abbreviation has become one of the most common methods of new word formation [355]. Sali Tagliamonte and Derek Denis [326] studied the IM techniques and behaviours of over 70 Toronto teenagers and found that IM is actually closer to a written version of normal conversation than to writing a letter or email.

Acronyms are also used extensively in online chat conversations. These short forms can be used to save time and to communicate specific meanings. For example, the abbreviation POS (*Parent Over Shoulder*) might be used to inform the chat buddy that the parent just walked in. Traditionally, acronyms and abbreviations are shortened versions of long compound words, technical terms or titles. For example, WWW stands for ‘World Wide Web’.

This study shows that such shortened word forms in text messages also occur in human-machine conversation in MSN Messenger. Since most AINI’s buddies are not concerned about grammatical conventions when typing instant messages, they use abbreviations, acronyms and phonetic substitutions extensively. Examples of such phenomenon are: “*How are you doing?*” becomes “*how u doin?*”, and “*before*” is replaced by “*b4*”. This helps to speed up the typed conversation among the buddies online.

**Table 7.2: Short forms used in IM Conversation Agents**

Shorthand	Expression	Freq	Shorthand	Expression	Freq
U/u	You	43	ru	are you	17
k	Ok	37	coz	because	17
btw	by the way	35	nt	no thanks	15
Ic	I see	29	oic	oh I see	13
bb	bye bye	28	B4	before	12
bfm	bye for now	25	plz	please	10
r	Are	24	dunno	Don’t know	9
HT	Hi there	22	s	see	8
msg	Message	20	wc	welcome	8
CU	see you	19	Y?	why	7
gotta	got to	18	wanna	want to	6
lol	Laughing out loud	18	<i>unidentified</i>		18

From the Chatlog, a wide variety of acronyms and abbreviations have been used. 422 unique short forms were found in the 3,280 messages. *U/u* (“you”) was the most commonly used in the IM conversation system. Table 7.2 shows a sample of the 422 short forms observed and



recorded during the study. 18 were labelled as “*unidentified*” because the meanings of them were unknown. These results differ from the OmniPod survey [356] based on IM use in the business environment. This is because this study was not restricted to a workplace environment, but rather based on open discussions in an unrestricted domain.

It was observed that the majority of the expressions in the logs are not acronyms, but rather shorthand. An example of words being shortened to emulate their spoken forms in the US accent includes the suffixation of ‘a’ to replace the prepositional ‘to’/‘of’ or ‘you’ is shown in the following example:

Want to à wanna  
got to à gotta  
Going to à gonna  
I’ve got you (I understand) à gotcha  
Sort of à sorta

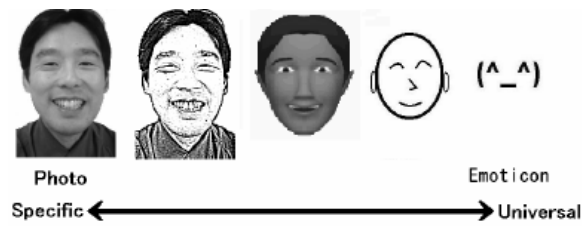
Interestingly, both the US accented “wanna” and “gotta” were also found in the human-machine communication. It does, however, stress that they are ‘non-standard and should generally be avoided in both speech and writing. This is indicative of the informality of their online speech – “they are chatting”.

In short, the data shows that paralinguistic abbreviations serve everyday needs, not just for human-human conversation but also appear in human-machine conversation. It could be said that the use of paralinguistic cues adds richness to the medium of communication. Most IM human users use shortened forms in their initial communications, and then change to more formal typed messages after a while when they realised that AINI appeared not able to understand the questions. Although AINI’s knowledge bases have limited abbreviations from the domain-specific and AAA knowledge bases, such as SARS, H5N1, AIDS, IT (*information technology*), USA (*United of America*), NY (*New York*), CU (*see you*), TTYL (*talk to you later*), etc., these shortened forms are not the regular forms used in the IM

conversation. Hence, in the future, it is proposed that paralinguistic properties should be included in CA knowledge bases to provide more human-like interaction.

### **7.2.3 Facial Expressions with Emoticons or Smileys**











Non-verbal expressions, such as emoticons or smileys, play another important role since the birth of CMC. An emoticon is a little picture created from plain text while larger pictures created from plain text are known as ASCII Art. An emoticon is a sequence of ordinary printable characters, such as ":-)", "^\_^", ".\_.", "XD", "X8", "-\_-", "=D", "=p", etc. or a small image, intended to represent a human facial expression and convey an emotion. In order to compensate for the lack of the parallel communicative channel provided by face-to-face interaction, IM's users have, over the course of time, developed and widely adopted what Werry [357] terms “orthographic strategies” to provide mechanisms of reflecting such elements as marked facial expressions and intonation. Studies on emoticons report positive effects on CMC [305, 324, 339, 358]. Those studies indicate that emoticons improve user experiences and interactions among participants and build enthusiasm toward participation and friendliness in intercultural communication [338], [339]. Emoticons are a form of paralinguistic cues commonly used as extended interpunction symbols in e-mail, IM, and online chat. In the case of a smile “:-)”, the colon “:” represents the eyes, the dash “-” represents the nose, and the right parenthesis “)” represents the mouth. To express emotions or facial expressions, there is an established IM ‘lingo’ of ‘smileys’ that are universally understood and used by the conversation agents. An example of the presentation of an emoticon is shown in Figure 7.4.

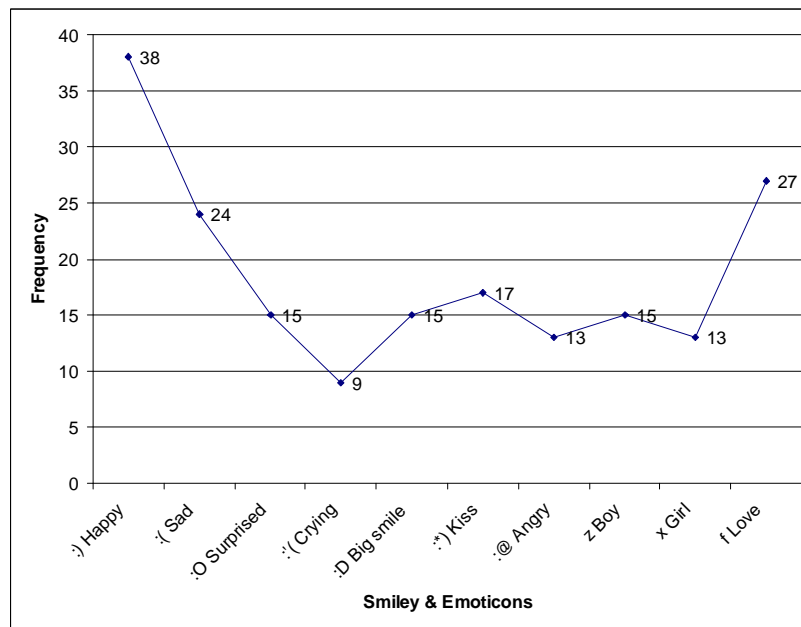


**Figure 7.4: Universal expression from Human to Computer Communication** by Scott McCloud [359]

Emoticons :) or :-) are symbols that represent happiness. Likewise, the symbols :( or :-( represent unhappiness. However, IM users have created their own emoticons to convey other feelings such as surprise, confusion, and anger to enrich their communications. There are different combinations that have been used during the conversation as shown in the Table 7.3 and Figure 7.5.

**Table 7.3: Facial expressions with Emoticons or Smileys used in IM Conversation Agents**

Smiley	Emoticon	Meaning	Freq
:)		Happy	38
:(		Sad	24
:O		Surprised	15
:'(		Crying	9
:D		Big smile	15
:*)		Kiss	17
:@		Angry	13
Z		Boy	15
X		Girl	13
F		Love	27



**Figure 7.5: Frequency of Smiley and Emoticons used in the IM Conversation Agents**

The excessive use of emoticons by AINI's buddies could cause some potential problems. First, MSN limits the number of emoticon images per IM. Second, the readability of sentences based on too little words and too many images turns out to be reduced. However, since emoticons offer shortcuts in communication, it is almost certain that a few will be included in the IM messages.

On the other hand, it was surprising to note that only a few emoticons appeared in the entire session (see visualisation expression in Figure 7.3). Out of the 186 emoticons used, 38 of them (20%) were "smileys". Moreover, a small number of IM users were responsible for using the majority of the 186 emoticons. This pattern is consistent with Walther and D'Addario's [121] finding. There are also users who like to emphasise the emotion in their queries by duplicating smiley's mouth, for example:

: ) ) ) ) ) )    à hearty laughter  
 : ( ( ( ( ( ( (    à deep sadness.

### 7.2.4 ASCII Arts

As the network bandwidth increases, many text-based IM clients provide a full range of emotional modifiers from tiny emoticons to enormous animations across the entire chat space. Excessive use of these features may ruin the simplicity of text. One of the major visual design issues is concerned with text-based interface. It is related to the question of how to maintain a consistent and coherent typographic layout, yet at the same time to provide a secondary channel for communicating paralinguistic cues. Figure 7.6 shows examples of the paralinguistic cues (ASCII arts) used by one of the AINI's buddies (user ID U1049) outweighing the meaning of words used in the corpus. In utterances #1 and #7, the ASCII art represents “love” and utterances #5 and #11 represents “health”.

#	ID	Utterance
<i>Continue from previous conversation.</i>		
1	U1049	( <u>  </u> ) (**) /-----\ /        *   ---   ~~   ~~
2	AINI	Is that a sign?
3	U1049	How u think?
4	AINI	My mind utilizes a sophisticated computer algorithm.
5	U1049	/ o) \ (o /
6	AINI	Oh.
7	U1049	.-. .-. .-. .-. .-. (( <u>  </u> )) (( <u>  </u> )) (( <u>  </u> )) (( <u>  </u> )) (( <u>  </u> )) '_ ' ' '_ ' '_ ' '_ ' '_ ' '_
8	AINI	I don't think very many people do that.
9	U1049	Are u sure
10	AINI	I am as sure as I can be.
11	U1049	,_O -_-< (*)/'(*)
12	AINI	I only hear that type of response less than five percent of the time
<i>Conversation continues...</i>		

**Figure 7.6: ASCII Arts used in IM Conversation Agents**

ASCII Arts consist of larger pictures drawn with plain text. Such pictures do not have to be limited to just ASCII characters. These "hand-crafted" ASCII sequences were not computer generated. The users require extra effort on each word to modify its typography in order to convey complex emotions even in a short sentence as shown in the example. This typographic approach attempts to convey meaning or message which may not be easily achieved by the simple emoticons. In this study, surprisingly only one of AINI's buddies, U1049, tried to use this approach to converse with AINI. Distinctive ASCII arts in the conversation to convey "love" or "health" were sent to AINI and shown in Figure 7.6.

Since AINI's knowledge bases are not currently equipped to interpret ASCII Arts, AINI normally will generate dynamic responses to queries using this type of typographic approach. Typical responses are *"is that a sign"*, *"I don't think very many people do that"*, *"oh"* and *"I only hear that type of response less than five percent of the time"*. As a result of this, AINI's buddies may become annoyed and irritated as they may perceive AINI to be "rude or stupid". Although AINI can learn new expressions in the future, it is almost impossible for CAs to recognise all "hand-crafted" expressions. This illustrates a major deficiency of CA's in their inability to interpret meanings from visual or graphical expressions such as ASCII Arts.

#### **7.2.5 Prosody Markers with Pauses and Voice Pitch**

Prosody refers to aspects of speech such as tone of voice, or inordinately slow or fast speech. Punctuation is one type of indicator of prosody in text. However, instant messages are characterised by the absence of punctuation as compared to formal written text documents [360]. In view of the similarity between IM and spoken communication, prosody can also play an important role in IM to help receivers interpret incoming messages.



### **7.3 Summary**

In this chapter, a study on the use of paralinguistic cues in AINI's IM exchanges with 65 buddies was carried out. Based on the corpus collected, it was observed that the usage varied between different users, especially with regard to paralinguistic properties. It was noted that IM human users at the beginning of the conversation, appeared to treat AINI as human buddy and they expressed their emotions through various forms of paralinguistic cues.

It is observed that CAs' messages in IM are more machine-like than IM conversations between human users. Human buddies frequently employ paralinguistic cues and they tend to use more intonations, abbreviations, facial expressions, ASCII Arts, and prosody markers than machines. Finally, in comparing IM transcripts between humans and machines, it is learnt that IM involving a CA differs significantly from human-human IM conversation. It was an objective of this thesis to develop the practical framework CAs which will be able to intimately understand conversations but also closely mimics a human conversation. In order to achieve this goal, the performance of practical CAs, such paralinguistic features should be included into the CA's knowledge bases in the future as discussed in Chapter Eight.



## **CHAPTER 8**

### **CONCLUSIONS**

#### **8.1 Introduction**

This thesis has presented the investigation and proposal of a practical framework for the development of conversation agents (CAs). In the course of this study, research work has included literature review, assessment of the state of problems and, the development of a CA architecture called AINI to demonstrate and evaluate the proposed concept through real-time experiments. AINI's knowledge is comprised of both open-domain and domain-specific knowledge bases in order to provide "meaningful" interaction with the users. An automatic knowledge acquisition tool (AKEA) has also been developed to assist in acquiring information in order to build the knowledge bases of the CA. In addition, the research has proposed a Web Knowledge Trust Model (WKTM) in order to address the issues on the trustworthiness of Internet resources. The study then examined the implications of linguistic features and paralinguistic cues used by the IM human users and AINI. This chapter presents a summary of the study, its contributions and the limitations of the research. This is followed by a discussion on the directions for future work.

##### **8.1.1 Summary**

This study has focused on the problem concerning the development of a practical framework for CAs. The research in this study has provided a better understanding of the system requirements and the development of a systematic approach for the construction of intelligent CAs based on agent architecture using a modular *N*-tiered approach.

A number of applications and modules have been presented to achieve the research objectives of developing a practical framework and improving the interoperability of the CA architecture through modular design as illustrated in the AINI framework [112, 113, 117]. The Natural Language Understanding and Reasoning (NLUR) module for example was developed to undertake full sentence parsing to produce grammatical categories and grammatical relationships by extracting pertinent information through the agent's knowledge, such as Noun Phrases and Verb Phrases as shown in the Appendix B. This module is extensible through polymorphic inheritance of components and was created based on a modular design.

A multilevel natural language query approach utilising multimodal interface and domain knowledge matrix model was also used. In addition to just providing an answer, the above approach aims to provide an ability to better understand the conversations and to mimic a human conversation while prolonging the exchange between the participants. A Domain Knowledge Matrix Model (DKMM) with multiple domain knowledge bases in the CA's architecture aims to enable the CA to provide the most appropriate response to the users.

The proposed Web Knowledge Trust Model (WKTM) and Automated Knowledge Extraction Agent (AKEA) are used to ensure the trustworthiness of the websites used in the development of the CA's domain-specific knowledge bases. By providing a flexible knowledge management application module in AKEA, this forms a useful device or tool for automatic acquisition of knowledge from different domains. The study demonstrates that the use of WKTM and AKEA improves the process of knowledge base development from existing online documents, as well as establishing the trustworthiness of the information sources.

In order to assess the proposed framework, relevant tools have been developed and an evaluation of their effectiveness has been carried out to validate the performance of the system. An interactive chat visualisation system called VisualChat tool for example has been developed and used for text analysis as described in Chapters 6 and 7. VisualChat can be used to visualise and analyse the human-machine conversation logs. VisualChat is capable of displaying the timeline of several textual conversations simultaneously and enables the scrutiny of utterance lengths and specific recurring keywords. VisualChat is also able to generate a graphical display that allows comparisons between the features of human and machine conversations.

In the evaluation of the CAs, both laboratory and public experiments with online users in real-time have been carried out. The results have shown that the proposed system is effective in engaging the users and in attempting to provide the most relevant answers instead of random replies. The study observed that CAs and human users share common properties in the use of linguistic features and paralinguistic cues. These human-computer interactions have been analysed and the findings contribute to the understanding of interaction between human and CAs. Results from the analysis will be useful for future development of CAs by utilising the commonalities found in this research.

### **8.1.2 Contributions**

The contributions made by this study during the course of design, implementation, evaluation and analysis of the proposed framework are described as follows:

- Presentation of a novel approach toward the development of practical CAs that allows users to communicate efficiently, flexibly and reliably. The framework developed

demonstrates an effective implementation and an exploration of the new paradigm of Computer Mediated Conversation (CMC) through CAs with an ability to understand conversations and closely mimics a human conversation.

- Presentation of the proposed framework and its ability to re-use and encapsulate expertise such as domain knowledge, natural language query and human-computer interface through plug-in components. A modified *N*-tiered architecture with additional channel service and domain service tiers contributes towards the goal of developing practical CA frameworks. Such architecture combines the advantages with simplicity, scalability, robustness and customisation.
- Presentation of the anthropomorphic modular design in the CA framework that allows users to communicate more effectively through multi channel communication interface including web, mobile devices and other network services.
- Presentation of a unique platform to enable the experimentation with, and analysis of the relationship between language and behaviour in the context of human-machine conversation. Results shown that top-down natural language query approach was a goal-driven approach similar to the way that human processes their language. The system utilises a multilevel natural language query approach and it has a higher likelihood of providing a better answer than systems which use a single strategy. This multilevel natural language query consists of Spell Check, Natural Language Understanding and Reasoning (NLUR); FAQChat; Index Search; Pattern Matching and Case Based Reasoning; and Supervised Learning by Domain Experts.

- Presentation of the Domain Knowledge Matrix Model (DKMM) in CA architecture which is capable of encompassing both open domain and domain specific knowledge. This will facilitate customisation of CAs and also empower the CAs to generate reliable and relevant responses and continue to prolong the conversation.
- Presentation of the Web Knowledge Trust Model (WKTM) for identifying reputable, credible, reliable and accountable websites. This novel approach contributes toward the building of CAs domain-specific knowledge from trusted sources. In spite of the use of semi-structured data from the World Wide Web (WWW) as the source for knowledge, the proposed model achieved an acceptable degree of confidence identifying and matching knowledge on the Web.

## 8.2 Limitations

While the design and implementation of the proposed CA has moved closer towards the original goal of developing flexible and adaptable CAs, the research also revealed several underlying limitations. These limitations are categorised as *agent knowledge* (domain knowledge), *agent brain* (natural language query) and *agent body* (human-computer interface) issues. They are detailed in the following subsections.

### 8.2.1 Agent Knowledge Issues

The experience from the knowledge extraction for CAs using WKTM has shown that the present process is labour intensive. This limitation is caused by the need to define a set of seeds, extract the corpus, evaluate the corpus, and to determine the trustworthiness of the websites using the proposed quantitative and qualitative assessments. An option to allow for a fully automatic approach with the ability to determine the seeds from the extracted

knowledge becomes apparent. A proposed solution could be an attempt to identify the seeds automatically using a set of heuristics and term expansion methods based on widely used sources such as the Word-Net [362] [363] [364] instead of bootstrapping from unstructured text.

In addition to the above limitation, an analysis of the paralinguistic cues has shown that the excessive use of emoticons, smileys, intonation, abbreviations, ASCII Art, and prosody markers by the CA's buddies is a problem. This will cause problems in continuing the conversation because of the inability of the CA to understand the paralinguistic features. Hence, means by which to tackle these paralinguistic properties should be included in CAs' knowledge bases in the future.

### **8.2.2 Agent Brain Issues**

Another limitation of the CA system is its lack of true support of anaphora and context in the agent brain. Anaphora is references made to previously-named objects. In English, pronouns like "I", "me", "you", "he", "she", "it", and so on usually refer to previously named people or things. Most of the time, CAs will ask the user's name, and incorporate this into future responses. Some CAs also attempts to identify the "topic" of conversation. Simply remembering the "name" or any other piece of information and storing it in a variable for later use does not constitute "understanding". The CA does not have any idea of what is going on in any sophisticated sense, nor it has any basis for "reasoning" the words. Pursuing a dialogue over a few exchanges will show that a CA really has no clue what it is talking about? Some CA developers pretend that their software can really identify the target of an anaphoric reference, but most such claims are just keyword-based. With the incorporation of additional database tables and intelligent techniques on Natural language understanding, it

should be possible to give CAs long-term or “permanent” memory through identification and personalisation of the CAs’ buddies.

### **8.2.3 Agent Body Issues**

Affect [365] and emotional intelligence [366] play crucial roles in human-computer interface, and are thus important components of CAs. A face with many expressions may lead to greater engagement, believability and amusement, and the CA could also command more attention [130]. Currently, AINI’s multimodal interface in the WebChat module is only supported by synthesis of prosodic and lip-sync speech, and co-verbal gestures by embodiment of avatar or anthropomorphism. Unfortunately, AINI lacks emotional facial expressions due to computational constraints. A solution to this limitation in order to fulfill the desired communicative function and to express the current emotional state within the conversation is the use of Emotion Recognition Module (ERM). They could be incorporated into CA’s framework. The ERM can be implemented by adhering to the XML interfaces at the channel service tier in the CA’s architecture and by including appropriate tags in the original input.

## **8.3 Future Work and Directions**

There are many possible directions in which this research could be continued in the future. This study has raised a couple of fundamental questions that are worth further investigation. As stated in Chapter 3, the proposed CA framework was designed using modified *N*-tiered architecture which shares common software services that could be accessed from any user interface, domain or task. Therefore, the CA could be supported on the web as well as mobile services, or other network services. The continual development of CAs could lead to new domain applications and can be applied to new and emerging domains and tasks.

Today, research on CAs has evolved from software robots to conversation agents existing in desktop computers or handheld devices. Another potential domain application might be “companion conversation robot (CCR)” [367-371] or “social robot” [340] an autonomous mobile robot [372-374] that can interact and communicate with humans. This application requires a conversation software agent equipped with a more complex decision making process and embedded with speech recognition technology. This type of CCR may one day act as a ‘partner’ to its users, offering assistance and companionship in domains of education, health care, household work, entertainment, and many others [375].

It is the author’s hope that this study will contribute towards the continual development of a practical and human-like CA for the benefit of the community. It is believed that this work has taken some small steps towards this ultimate goal.



## REFERENCES

- [1] R. Kurzweil, *The Age of Intelligent Machines*. Cambridge, MA: MIT Press, 1990.
- [2] J. Haugland, *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press, 1985.
- [3] R. J. Schalkoff, *Artificial Intelligence: An Engineering Approach*. New York: McGraw-Hill, 1990.
- [4] E. Charniak and D. McDermott, *Introduction to Artificial Intelligence*. Reading, MA: Addison-Wesley, 1985.
- [5] M. L. Mauldin, "Chatterbots, Tinymuds, and the Turing Test: Entering the Loebner Prize Competition," in *AAAI 94*, 1994.
- [6] E. Rich and K. Knight, *Artificial Intelligence (2nd Edition)*. New York: McGraw-Hill, 1991.
- [7] P. H. Winston, *Artificial Intelligence (3rd Edition)*. Reading, MA: Addison-Wesley, 1992.
- [8] R. E. Bellman, *An Introduction to Artificial Intelligence: Can Computers Think?* San Francisco: Boyd & Fraser, 1978.
- [9] G. F. Luger and W. A. Stubblefield, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving (2nd Edition)*. Redwood City, CA: Benjamin/Cummings, 1993.
- [10] T. Bickmore and R. W. Picard, "Establishing and Maintaining Long-Term Human-Computer Relationships," *Trans. on Computer-Human Interaction*, vol. 12, pp. 293 - 327, 2004.
- [11] B. Reeves and C. Nass, "The media equation: how people treat computers, television, and the new media like real people and places," Center for the Study of Language and Information, Cambridge University Press, Stanford, CA 1996.
- [12] C. Creed and R. Beale, "Agent Abuse: The Potential Dangers of Socially Intelligent Embodied Agents," in *CHI 2006*, Montreal, Canada., 2006.
- [13] A.M.Turing, "Computing Machinery and Intelligence," *MIND the Journal of the Mind Association*, vol. LIX, pp. 433-460, 1950.
- [14] B. Gates, "A Revolution in Communications," The Empire Club foundation, 20 July 2007, <http://www.empireclubfoundation.com/details.asp?FT=yes&SpeechID=747>.
- [15] N. Block, "Psychologism and behaviorism," *The Philosophical Review*, vol. XC, pp. 5-43, 1981.
- [16] R. French, "Subcognition and the limits of the Turing test," *Mind*, vol. 99, pp. 53-65, 1990.
- [17] J. Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, vol. 3, pp. 417-57, 1980.
- [18] S. M. Shieber, "Does the Turing Test Demonstrate Intelligence or Not?," in *Twenty-First National Conference on Artificial Intelligence (AAAI-06)*, Boston, MA, 2006.
- [19] Human-Evolution, "What is Artificial Intelligence (AI)?," Center for Future Human Evolution, 22 July 2007, <http://www.human-evolution.org/aiwhat.php>.
- [20] F. Rangel, "Turing Test Proves 2-Year-Olds Not Human," 20 July 2007, <http://www.bbspot.com/News/2006/10/turing-test.html>.
- [21] P. Cohen, "If Not Turing's Test, Then What?," in *AI Magazine*. vol. 26, 2006.
- [22] P. Hayes and K. Ford, "Turing test considered harmful," in *International Joint Conference on Artificial Intelligence*, Montreal, Quebec, 1995, pp. 972-997.
- [23] H. G. Loebner, "Home page of the Loebner prize—the first Turing test," Loebner Prize, 25 July 2007, <http://www.loebner.net/Prize/loebner-prize.html>.

- [24] W. J. Rapaport, "The Turing Test," in *Encyclopedia of Language and Linguistics* (2nd Edition) Buffalo: Elsevier, forthcoming.
- [25] S. M. Shieber, "Lessons from a restricted Turing Test," *Communications of the Association for Computing Machinery*, vol. 37, pp. 70-8, 1994.
- [26] S. M. Shieber, "On Loebner's lessons," *Communications of the ACM*, vol. 37, pp. 83 - 84, 1994.
- [27] H. G. Loebner, "In Response [to Shieber 1994a]," *Communications of the ACM*, vol. 37, pp. 79-82, 1994.
- [28] G. Oppy and D. Dowe, "The Turing Test," *Stanford Encyclopedia of Philosophy*, 15 July 2007, <http://plato.stanford.edu/entries/turing-test/>.
- [29] P. Ross, "Artificial Intelligence at Edinburgh University," Edinburgh University, 23 July 2007, <http://www.inf.ed.ac.uk/about/AIhistory.html>.
- [30] D. B. Lenat and G. R. V, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*: Addison-Wesley, 1989.
- [31] TriumphPC, "The John Lennon Artificial Intelligence Project," 20 Jun 2007, <http://triumphpc.com/john-lennon-project/index2.shtml>.
- [32] J. Weizenbaum, "ELIZA - A computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, pp. 36-45, 1966.
- [33] K. M. Colby, *Artificial Paranoia: A Computer Simulation of Paranoid Processes*: Pergamon Press Inc., 1975.
- [34] K. M. Colby, "Modeling a Paranoid Mind," *Behavioral and Brain Sciences*, vol. 4, pp. 515-560, 1981.
- [35] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, pp. 36-45, 1966.
- [36] R. F. Simmons, "Computational Linguistic, Natural Language Question Answering System: 1969," *Communications of the ACM*, vol. XIII, pp. 15-29, 1970.
- [37] J. Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation*. New York, NY, USA: W. H. Freeman & Co, 1976.
- [38] R. S. Wallace, H. Tomabechi, and D. Aimless, "Chatterbots Go Native: Considerations for an eco-system fostering the development of artificial life forms in a human world," Pandorabots, 20 July 2007, <http://www.pandorabots.com/pandora/pics/chatterbotsgonative.doc>.
- [39] G. Güzeldere and S. Franchi, "Dialogues with Colorful Personalities of Early AI," *Stanford Electronic Humanities Review*, vol. 4, 1995.
- [40] BBC, "Alice chatbot wins for third time," BBC News, 10 February 2007, <http://news.bbc.co.uk/1/hi/technology/3672424.stm>.
- [41] R. S. Wallace, "The Anatomy of A.L.I.C.E.," A.L.I.C.E. Artificial Intelligence Foundation, Inc., 20 July 2007, <http://www.alicebot.org/anatomy.html>.
- [42] G. K. Zipf, *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison- Wesley., 1949.
- [43] Chatterboxchallenge, "ALICE Winner of Chatterbox Challenge 2004," Chatterbox Challenge, 22 July 2007, <http://www.chatterboxchallenge.com/>.
- [44] C. McKinstry, "Jackie and the Brain," Kuro5hin, 20 July 2007, <http://www.kuro5hin.org/story/2005/6/11/224459/073>.
- [45] MindPixel, "GAC-80K Mindpixel," 10 July 2006, <http://sluggish.uni.cc/mindpixel/gac80k-06-july-2005.html>.
- [46] A. Powers, "An Easy To Use Dialogue System: AIMLE," Carnegie Mellon University, Pittsburgh, PA. This is an unpublished technical report on improvements to a basic dialogue system, *AIML*. It's available upon request, unpublished.

- [47] A. Powers, "Improving Pattern-Matching For Better Language Processing," 20 July 2007,  
<http://www.contrib.andrew.cmu.edu/~apowers/articles/PatternMatchingImprovements.pdf>.
- [48] D. Aimless, "Pandorawriter - Converts Dialog to AIML Categories," 20 July 2007,  
<http://www.pandorabots.com/pandora/aiml-converter-intro.html>.
- [49] Pandorabots, "Pandorabots is the place where you can create and unleash virtual personalities," Pandorabots.com, 22 Jul 2007, <http://www.pandorabots.com/>.
- [50] D. Maass, "Bots Will Be Bots," maassive.com, 20 July 2007,  
[http://maassive.com/?page\\_id=203](http://maassive.com/?page_id=203).
- [51] R. S. Wallace, "The Annotated A.L.I.C.E. AIML," ALICE Foundation, 22 July 2007,  
<http://www.alicebot.org/aiml/aaa/>.
- [52] R. S. Wallace, "ALICE and AIML for ESL," Alice Foundation, 22 July 2007,  
<http://www.alicebot.org/ESL.html>.
- [53] H. Shah and V. Pavlika, "Text based Dialogical E-Query Systems: gimmick or convenience?," in *10th International Conference on Speech and Computer*, 2005, pp. 425-428.
- [54] F. Zambetta, G. Catucci, F. Abbattista, and G. Semeraro, "SAMIR: Your 3D Virtual Bookseller," in *Knowledge-Based Intelligent Information and Engineering Systems*. vol. 2773 Berlin: Springer, 2003.
- [55] E. Figa and P. Tarau, "The VISTA Project: An Agent Architecture for Virtual Interactive Storytelling," *SIGGROUP Bulletin*, vol. 23, 2002.
- [56] A. Egges, G. Papagiannakis, and N. Magnenat-Thalmann, "An Interactive Mixed Reality Framework for Virtual Humans," MiraLab, 20 July 2007,  
<http://www.miralab.unige.ch/papers/403.pdf>.
- [57] B. Abu Shawar, E. Atwell, and A. Roberts, "FAQChat as an Information Retrieval System," in *2nd Language and Technology Conference*, Wydawnictwo Poznanskie, Poznan, Poland, 2005, pp. 274-278.
- [58] M. Gorseta, I. Cipar, and D. Mikic, "Experimental Performance Evaluation of Networked Virtual Reality ?Services in UMTS Network," in *IEEE MELECO 2004*, Dubrovnik, Croatia, 2004.
- [59] A. Santangelo, A. G. A. Augello, G. Pilato, and S. Gaglio, "A Chat-Bot based Multimodal Virtual Guide for Cultural Heritage Tours," in *The 2006 World Congress in Computer Science Computer Engineering, and Applied Computing*, Las Vegas, Nevada, 2006.
- [60] G. Weber and G. Gibbs, "Emile: Using a chatbot conversation to enhance the learning of social theory," University of Huddersfield, Huddersfield 05/S/01, 2001.
- [61] R. Moore and G. Gibbs, "Emile: Using a chatbot conversation to enhance the learning of social theory," University of Huddersfield, Huddersfield, England 2002.
- [62] L. Fryer and R. Carpenter, "Emerging Technologies: Bots as Language Learning Tools," *Language Learning & Technology*, vol. 10, pp. 8-14, 2006.
- [63] M. Ginsburg and S. Weisband, "Social Capital and Volunteerism in Virtual Communities: The Case of the Internet Chess Club," in *35th Hawaii International Conference on System Sciences*, 2002.
- [64] K. Brennan, "The Managed Teacher: Emotional labour, Education and Technology," *Educational Insights*, vol. 10, 2006.
- [65] O. D. Pietro, M. D. Rose, and G. Frontera, "Automatic Update of AIML Knowledge Base in E-Learning Environment," in *Computers and Advanced Technology in Education*, 2005.
- [66] F. Meijerink, P. Maanen, P. v. Maanen, A. J. v. Vliet, and A. Nijholt, "Disclosure with an Emotional Intelligent Synthetic Partner ,," in *Workshop: Tools for Psychological*

- Support during Exploration Missions to Mars and Moon, European Space Research and Technology Centre (ESTEC), Noordwijk, The Netherlands, 2007.*
- [67] S. Kshirsagar and N. Magnenat-Thalmann, "Virtual Humans Personified," in *Autonomous Agents Conference (AAMAS) 2002*, 2002, pp. 356-359.
  - [68] A. M. Galvao, F. A. Barros, A. M. M. Neves, and G. L. Ramalho, "Persona-AIML: an architecture for developing chatterbots with personality," in *Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004)*, 2004, pp. 1266 - 1267.
  - [69] H. Tanaka, T. Tokunaga, Y. Shinyama, and H. Prendinger, "Animated Agents Capable of Understanding Natural Language and Performing Actions," in *Life-Like Characters*, H. Prendinger and M. Ishizuka, Eds. Heidelberg: Springer, 2004, pp. 429-443.
  - [70] D. Feng, E. Shaw, J. Kim, and E. Hovy, "An intelligent discussion-bot for answering student queries in threaded discussions," in *International Conference on Intelligent User Interfaces*, Sydney, Australia, 2006.
  - [71] A. Kerlya, P. Hallb, and S. Bulla, "Bringing chatbots into education: Towards natural language negotiation of open learner models," *Knowledge-Based Systems*, vol. 20, pp. 177-185, 2007.
  - [72] E. Reategui, E. Boff, and J. A. Campbell, "Using virtual characters in personalized recommendations," in *38th SIGCSE technical symposium on Computer science education*, Covington, Kentucky, 2007.
  - [73] Newswire, "Buddhabot is an Inspirational AI that Advocates Quantum Philosophy to Millions," PRWeb, [http://www.buddhabots.com/Downloads/Press\\_Release.pdf](http://www.buddhabots.com/Downloads/Press_Release.pdf).
  - [74] G. Weber, "Data representation and algorithms for biomedical informatics applications," in *Division of Engineering and Applied Sciences* Cambridge, Massachusetts: Harvard University, 2005.
  - [75] G. Weber, L. Ohno-Machado, and S. Shieber, "Representation in stochastic search for phylogenetic tree reconstruction," *Journal of Biomedical Informatics*, vol. 39, pp. 43-50, 2006.
  - [76] B. P. Kerfoot, H. J. Baker, T. L. Jackson, W. C. Hulbert, D. D. Federman, R. D. Oates, and W. C. DeWolf, "A multi-institutional randomized controlled trial of adjuvant Web-based teaching to medical students.," *Academic Medicine*, vol. 81, pp. 224-230, 2006.
  - [77] AAMC, "Virtual Patients Reference Center," MedEdPORTAL, 20 July 2007, <http://www.aamc.org/meded/mededportal/vp>.
  - [78] C. Torrey, A. Powers, S. Fussell, and S. Kiesler, "Exploring Adaptive Dialogue Based on a Robot's Awareness of Human Gaze and Task Progress," in *Human-Robot Interaction (HRI'07)*, 2007.
  - [79] A. Powers, A. D. Kramer, S. Lim, J. Kuo, and S. Sau-lai LeeKiesler, "Eliciting information from people with a gendered humanoid robot," in *IEEE International Workshop on Robot and Human Interactive Communication*, Roman, 2005, pp. 158 - 163.
  - [80] S. Rao, S. Garg, H. Lieberman, and H. Liu, "Commonsense via Instant Messaging," Massachusetts Institute of Technology, 20 July 2007, [http://ocw.mit.edu/NR/temp/5C7CF160-84C0-4E56-BD32-5F36B6CCE2AE/0/proj\\_file9\\_wor.pdf](http://ocw.mit.edu/NR/temp/5C7CF160-84C0-4E56-BD32-5F36B6CCE2AE/0/proj_file9_wor.pdf).
  - [81] R. Schumaker, M. Ginsburg, H. Chen, and Y. Liu, "An Evaluation of the Chat and Knowledge Delivery Components of a Low-Level Dialog System: The AZ-ALICE Experiment," *Decision Support Systems*, vol. 42, pp. 2236-2246, 2007.
  - [82] R. Schumaker, Y. Liu, M. Ginsburg, and H. Chen, "Evaluating Mass Knowledge Acquisition using the ALICE Chatterbot," *International Journal of Human-Computer Studies*, vol. 64, pp. 1132-1140, 2006.

- [83] R. Schumaker, Y. Liu, M. Ginsburg, and H. Chen, "Evaluating the Efficacy of a Terrorism Question Answer System: The TARA Project," *Communications of the ACM*, vol. 50, pp. 74-80, 2007.
- [84] R. Schumaker and H. Chen, "Interaction Analysis of the ALICE Chatterbot: A Two Study Investigation of Dialog and Domain Questioning," *IEEE Systems Man and Cybernetics, Part A*, Review.
- [85] R. Carpenter, "About the Jabberwacky AI," Jabberwacky.com, 20 July 2007, <http://www.jabberwacky.com/j2about>.
- [86] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, second edition: Prentice Hall, 2003.
- [87] F. Roberts, "The A.I. of Elbot," Software Developer's Journal, 20 July 2007, <http://en.sdjournal.org/?module=products&moduleAction=articleInfo&value=36>.
- [88] S. M. Shieber, "Lessons from a Restricted Turing Test," Harvard University Technical Report TR-19-92, Sept. 1992, Revision 4.
- [89] M. L. Mauldin, "Chatterbots, Tinymuds, And The Turing Test: Entering The Loebner Prize Competition," in *AAAI-94*, 1994.
- [90] J. H. Moor, "An analysis of the Turing test," *Philosophical Studies*, vol. 30, pp. 249-257, 1976.
- [91] T. Winograd, *Understanding Natural Language*: Academic Press, 1972.
- [92] T. Winograd, "Procedures as a Representation for Data in a Computer Program for Understanding Natural Language," *Cognitive Psychology*, vol. 3, 1972.
- [93] W. Woods, "Progress in Natural Language Understanding: An Application to Lunar Geology," in *National Conference of the American Federation of Information Processing Societies*, 1973.
- [94] AskJeeves, "AskJeeves," IAC Search & Media, 22 July 2007, <http://www.ask.com/>.
- [95] Z. Zheng, "AnswerBus Question Answering System," in *Conference on Human Language Technology*, 2002.
- [96] J. Lin, V. Sinha, B. Katz, K. Bakshi, D. Quan, D. Huynh, and D. Karger, "What Makes a Good Answer? The Role of Context in Question Answering," in *9th International Conference on Human-Computer Interaction*, 2003.
- [97] D. B. Lenat, "CYC: A large-scale investment in knowledge infrastructure," *Communications of the ACM*, vol. 38, pp. 33-38, 1995.
- [98] K. Panton, C. Matuszek, D. Lenat, D. Schneider, M. Witbrock, N. Siegel, and B. Shepard, "Common Sense Reasoning – From Cyc to Intelligent Assistant," in *Ambient Intelligence in Everyday Life*, Y. Cai and J. Abascal, Eds. Berlin Heidelberg: Springer-Verlag, 2006.
- [99] P. Singh, T. Lin, E. T. Mueller, G. Lim, T. Perkins, and W. L. Zhu, "Open Mind Common Sense: Knowledge Acquisition from the General Public," in *On the Move to Meaningful Internet Systems, 2002 - DOA/CoopIS/ODBASE 2002*, London, UK, 2002, pp. 1223 - 1237.
- [100] M. Richardson and P. Domingos, "Building Large Knowledge Bases by Mass Collaboration," in *KCAP'03*, Sanibel Island, Florida, USA, 2003.
- [101] C. McKinsty, R. Dale, and M. J. Spivey, "Action dynamics reveal parallel competition in decision making," *Psychological Science*, vol. 19, pp. 22-24, 2008.
- [102] ALICE, "ALICE Silver Edition," ALICE Foundation, 20 August 2007, <http://www.alicebot.org/join.html>.
- [103] J. Rickel and W. L. Johnson, "Animated agents for procedural training in virtual reality: Perception, cognition and motor control.," *Applied Artificial Intelligence Journal*, vol. 13, pp. 343-382, 1998.
- [104] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, *Embodied Conversational Agents*. Cambridge, MA.: MIT Press, 2000.



- [105] C. M. Karat, J. Vergo, and D. Nahamoo, "Conversational interface technologies," in *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, J. Jacko and A. Sears, Eds. Mahwah, NJ: Lawrence Erlbaum Assoc., 2003, pp. 169-186.
- [106] J. Lai and N. Yanlelovich, "Conversational speech interfaces," in *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, J. Jacko and A. Sears, Eds. Mahwah, NJ: Lawrence Erlbaum Assoc., 2003, pp. 698-713.
- [107] S. L. Oviatt, C. Darves, and R. Coulston, "Toward Adaptive Conversational Interfaces: Modelling Speech Convergence with Animated Personas," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 11, September 2004.
- [108] M. Halpern, "The Trouble with the Turing Test," *A Journal of Technology and Society*, vol. 11, pp. 42-63, 2006.
- [109] S. M. Shieber, "Lessons from a Restricted Turing Test," *Communications of the ACM*, vol. 37, pp. 70-78, 1994.
- [110] O. S. Goh, N. Aida, and M. Shamsuri, "Managing Communication During Crisis in the Internet Using Software Robot," in *National Communication Seminar II 2004*, Langkawi, Malaysia, 2004, pp. 77 - 84.
- [111] O. S. Goh, "A Global Internet-based Crisis Communication: A Case study on SARS using Intelligent Agent," University Technical Malaysia Melaka, Malacca, Malaysia, Technical Report PJP/2003/FTMK(1), 2005.
- [112] O. S. Goh, C. C. Fung, and M. P. Lee, "Intelligent Agents for an Internet-based Global Crisis Communication System," *Journal of Technology Management and Entrepreneurship*, vol. 2, pp. 65-78, 2005.
- [113] O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "Embodied Conversational Agents for H5N1 Pandemic Crisis," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 11(2), pp. 282-288, 2007.
- [114] K. Barrett, D. LaPointe, and K. Greysen, "Speak2Me: using synchronous audio for ESL teaching in Taiwan," Centre for Distance Education, Athabasca University, Online Software Evaluation Report: R28/0401, 2004.
- [115] S. Rupley, "The Human Computer - Want to make a friend online?," Oddcast Inc, 20 July 2007, <http://www.oddcast.com/home/index.php?id=117>.
- [116] P. Millican, "Elizabeth Conversation Program documentation," Leeds Electronic Text Centre, 20 July 2007, <http://www.etext.leeds.ac.uk/elizabeth/>.
- [117] O. S. Goh, C. C. Fung, C. Ardil, K. W. Wong, and A. Depickere, "A Crisis Communication Network Based on Embodied Conversational Agents System with Mobile Services," *Journal of Information Technology*, vol. 3, pp. 257-266, 2006.
- [118] N. S. Baron, "Discourse Structures in Instant Messaging," in *Computer-Mediated Conversation.*, Susan Herring, Ed. Cresskill, NJ: Hampton Press, 2005.
- [119] J. Carey, "Paralanguage in computer mediated communication," in *18th Annual Meeting of the Association for Computational Linguistics*, University of Pennsylvania, Philadelphia, Pennsylvania, USA, 1980.
- [120] M. Lea and R. Spears, "Paralanguage and social perception in computer-mediated communication," *Journal of Organizational Computing*, vol. 2, pp. 321-341, 1992.
- [121] J. B. Walther and K. P. D'Addario, "The impacts of emoticons on message interpretation in computer-mediated communication," *Social Science Computer Review*, vol. 19, pp. 323-345, 2001.
- [122] Y.-o. Yum and K. Hara, "Computer-Mediated Relationship Development: A Cross-Cultural Comparison," *Journal of Computer-Mediated Communication*, vol. 11, pp. 133-152, 2006.

- [123] B. J. Poole , M. Axmann, C. M. Calongne, and D. Cox, "To Chat or to Chatter: Making Online Sense of the Chat Room Environment," in *The TCC 2003 Online Conference*, 2003.
- [124] ResearchandMarkets, "Global - Mobile Data - SMS & Instant Messaging," Research and Markets, 5 August 2007, [http://www.researchandmarkets.com/reports/446943/global\\_mobile\\_data\\_sms\\_and\\_instant\\_messaging.pdf](http://www.researchandmarkets.com/reports/446943/global_mobile_data_sms_and_instant_messaging.pdf).
- [125] J. Tang, N. Yankelovich, and J. Begole, "ConNexus to Awarenex: Extending Awareness to mobile Users," in *CHI 2001*, New York, 2001.
- [126] L. Palen, M. Salzman, and E. Youngs, "Going wireless: behavior & practice of new mobile phone users," in *the 2000 ACM conference on Computer Supported Cooperative Work*, Philadelphia, 2000, pp. 201-210.
- [127] M. Yuasa, K. Saito, and N. Mukawa, "Emoticons Convey Emotions without Cognition of Faces: An fMRI Study," in *CHI 2006*, Montréal, Québec, Canada, 2006.
- [128] R. Birdwhistell, "Essays on Body Motion Communication," in *Kinetics and Context*: University of Pennsylvania Press, 1970.
- [129] T. M. Holtgraves, S. J. Ross, C. R. Weywadt, and T. L. Han, "Perceiving artificial social agents," *Computers in Human Behavior*, vol. 23, pp. 2163–2174, 2007.
- [130] T. Koda and P. Maes, " Agents with faces: the effect of personification," in *5th IEEE International Workshop on Robot and Human Communication*, Tsukuba, Japan, 1996, pp. 189-194.
- [131] T. Wu, "Programming Language Design in Cybernetics Applications," in *IEEE International Conference on Systems, Man, and Cybernetics*, Taipei, 2006, pp. 2368-2373.
- [132] R. S. Wallace, "Don't Read Me: Program dB," 25 July 2007, <http://www.alicebot.org/articles/wallace/dont-dB.html>.
- [133] R. K. Dybvig, *The Scheme Programming Language*. Cambridge, Massachusetts: The MIT Press, 2003.
- [134] J. A. McCann and K. J. Manning, "Performance Management Tool for Interoperable Environments," in *4th Annual Conference of the British Computer Society Client/Server Group, CSG '97*, 1997, pp. 60-70.
- [135] M. M. Bray, "Middleware- Software Technology Roadmap," Software Engineering Institute (SEI), Carnegie Mellon University, 25 July 2007, <http://www.sei.cmu.edu/str/descriptions/middleware.html>.
- [136] A. Ramirez, "Three-Tier Architecture," Linux Journal, 30 July 2007, <http://www.linuxjournal.com/article/3508>.
- [137] G. Schussel, "Client/Server Past, Present, and Future," Digital Consulting, Inc., 25 July 2007, <http://www.dciexpo.com/geos/dbsejava.htm>.
- [138] W. W. Eckerson, "Three Tier Client/Server Architecture: Achieving Scalability, Performance, and Efficiency in Client Server Applications," *Open Information Systems*, vol. 3, 1995.
- [139] D. Sadoski, S. Comella-Dorda, P. Clements, and F. Rogers, "Three Tier Software Architectures - Software Technology Roadmap," Software Engineering Institute (SEI), Carnegie Mellon University, 25 July 2007.
- [140] R. P. Bonasso, R. J. Firby, E. Gat, D. Kortenkamp, D. Miller, and M. Slack, "Experiences with an architecture for intelligent, reactive agents.," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 9, 1997.
- [141] Corda, "Dashboard Design Strategies for Human Capital Management," Corda.com, 30 July 2007, [http://www.corda.com/company/resources/white\\_papers/HCM%20Architecture%20v1.0.pdf](http://www.corda.com/company/resources/white_papers/HCM%20Architecture%20v1.0.pdf).

- [142] J. Walton, R. E. Filman, and D. J. Korsmeyer, "The Evolution of the DARWIN System," in *ACM Symposium on Applied Computing*, Como, Italy, 2000, pp. 971-977.
- [143] P. P. Zhang, "Globally-consistent 3D simultaneous localization and mapping with multi-sensor fusion," in *Faculty of Computer Science* Halifax, NS, Canada.: Dalhousie University, 2007.
- [144] I. R. Nourbakhsh, K. Sycara, M. Koes, M. Yong, M. Lewis, and S. Burion, "Human-Robot Teaming for Search and Rescue," *Pervasive Computing*, vol. 4, pp. 72-78, 2005.
- [145] M. Montemerlo, N. Roy, and S. Thrun, "Perspectives on Standardization in Mobile Robot Programming : The Carnegie Mellon Navigation (CARMEN) Toolkit," 25 July 2007, <http://web.mit.edu/nickroy/www/papers/iros03.pdf>.
- [146] F. Z. Qureshi, D. Terzopoulos, and R. Gillette, "The Cognitive Controller: A Hybrid, Deliberative/Reactive Control Architecture for Autonomous Robots," in *Innovations in Applied Artificial Intelligence and Expert Systems*, Ottawa, Canada, 2004.
- [147] V. Kulyukin, "Human-Robot Interaction Through Gesture-Free Spoken Dialogue," *Autonomous Robots*, vol. 16, pp. 239-257, 2004.
- [148] D. Rautela, G. Markwardt, and A. Kamel, "Combining Open APIs(Parlay/JAIN) & Software Agents for Next Generation Mobile Services," in *1st International Workshop on M-Services - Concepts, Approaches, and Tools*, Lyon, France, 2002.
- [149] Y.-T. Lu, C.-W. Chang, and G.-D. Chen, "Using a Programmable Storytelling Robot to Motivate Learning Second Language," in *IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, 2007, pp. 841-844.
- [150] A. Caglaya, M. Snorrason, J. Jacoby, J. Mazzu, R. Jones, and K. Kumar, "Learn Sesame - a Learning Agent Engine," Opensesame, Cambridge, Massachusetts 1997.
- [151] C.-H. Lu, G.-F. Chiou, M.-Y. Day, and C.-S. Ong, "Using Instant Messaging to Provide an Intelligent Learning Environment," in *ITS 2006*, Berlin Heidelberg, 2006, pp. 575 – 583.
- [152] N. Bush, "Getting Started With Program D," ALICE Foundation, 30 July 2007, <http://www.alicebot.org/resources/programd/readme.html>.
- [153] T. Wu, "Constructing a Topological Space over a Relational Database System," in *International Conference on Computational Intelligence for Modeling, Control, and Automation (CIMCA )*, Gold Coast, 2004, pp. 80-86.
- [154] P. Rydell, "Program E: Documentation and downloads," Rydell.com, 25 July 2007, <http://www.rydell.com/programe>.
- [155] J. F. Allen, D. K. Byron, M. O. Dzikosvska, G. Fergusson, L. Galescu, and A. Stent, "Towards conversational Human-Computer Interaction," in *AI magazine*, 2001.
- [156] O. S. Goh and C. C. Fung, "Automated Knowledge Extraction from Internet for a Crisis Communication Portal," in *First International Conference on Natural Computation* Changsha, China: Lecture Notes in Computer Science (LNCS), 2005, pp. 1226-1235.
- [157] J. Cassell, "Embodied Conversation: Integrating Face and Gesture into Automatic Spoken Dialogue Systems," in *Spoken Dialogue Systems*, Luperfoy, Ed.: MIT Press, to appear.
- [158] J. Cassell, T. Bickmore, M. Billinghamurst, L. Campbell, K. Chang, and H. Vilhjalmsson, "Embodiment in conversational interfaces: Rea," in *the CHI 99 Conference on Human Factors in Computing Systems*, New York, 1999.
- [159] S. Yamada and N. Nagino, "Constructing a personal Web map with anytime-control of Web robots," *International Journal of Cooperative Information Systems*, vol. 11, pp. 1-19, 2002.
- [160] M. Huynh, "Brain, the Framework for a Chatterbot," Mikex Studios, 25 July 2007, <http://www.mikexstudios.com/research/Brain%20-%20Chatterbot%20Framework.pdf>.
- [161] P. Cosi, C. Drioli, F. Tesser, and G. Tisato, "INTERFACE toolkit: a new tool for building IVAs," in *Intelligent Virtual Agents Conference (IVA'05)*, KOS Greece, 2005.



- [162] C. S. Nass, J. Tauber, and R. Ellen, "Computers are Social Actors," in *CHI'94: Human Factors in Computing Systems*, Boston, MA, 1994, pp. 72-77.
- [163] Microsoft, "Microsoft Agent," Microsoft Corporation, 1 August 2007, <http://www.microsoft.com/msagent/default.asp>.
- [164] Adobe, "Adobe Flash Player," Adobe Systems Incorporated, 1 August 2007, <http://www.macromedia.com/software/flash/about/>.
- [165] Oddcast, "Oddcast Technology," Oddcast Inc, 1 August 2007, <http://www.oddcast.com>.
- [166] H. Sulaiman, "Software robot for PDAs," in *Computimes, News Strait Times Press* Malaysia, 2003.
- [167] AAAI, "AINI: AI software interactive robot gets recognition," American Association for Artificial Intelligence, 8 June 2007, <http://www.aaai.org/AITopics/newstopics/nlp1.html>.
- [168] HighBeam, "Software robot for PDAs," HighBeam Research, 2 May 2006, <http://www.highbeam.com/doc/1P1:84894246/Software+robot+for+PDAs.html>.
- [169] Microsoft, "Microsoft Windows Mobile," 1 May 2006, <http://www.microsoft.com/windowsmobile/>.
- [170] M. Mondok, "60% of IM users prefer MSN Messenger," comScore Networks, 1 August 2007, <http://arstechnica.com/journals/microsoft.ars/2006/4/11/3557>.
- [171] O. S. Goh, S. Sahib, and R. Elangsegaran, "An Intelligent Virtual Financial Advisor System (IVFAS)," in *2nd International Conference on Neural Network and Computational Intelligence (NCI 2004)*, Zurich, Switzerland, 2004, pp. 146-151.
- [172] N. Dahlbäck and A. Jönsson, "Integrating domain specific focusing in dialogue models," in *Eurospeech'97*, Rhodes, Greece, 1997, pp. 2215-2218.
- [173] O. S. Goh, A. Depickere, C. C. Fung, and K. W. Wong, "Domain Matrix Knowledge Model for Embodied Conversation Agents," in *5th International Conference on Research, Innovation & Vision for the Future (RIVF'07)*, Hanoi, Vietnam, 2007.
- [174] Marquis, *Who's Who is Science and Engineering 2008 - 2009*, 10th Anniversary ed. NJ, USA: Marquis Who's Who LLC, 2008.
- [175] O. S. Goh, C. C. Fung, and A. Depickere, "Domain Knowledge Query Conversational Bot in Instant Messaging (IM)," *Knowledge-Based Systems*, vol. 21, 2008.
- [176] E. T. Bell, "Mathematics: Queen and Servant of Science," in *MAA Spectrum Series*, 1987.
- [177] W. Saba, "The Structure of Commonsense Knowledge," in *14th Australian Joint Conference on Artificial Intelligence*, Adelaide, Australia, 2001.
- [178] P. Singh and M. Minsky, "A Ten Year Roadmap to Machines with Common Sense," Massachusetts Institute of Technology 2002.
- [179] NIST, "Text REtrieval Conference (TREC)," 11 October 2006, <http://trec.nist.gov/>.
- [180] H. Loebner, "Loebner Prize," 6 October 2006, <http://www.loebner.net/Prizetf/loebner-prize.html>.
- [181] Chatterboxchallenge, "ALICE Winner of Chatterbox Challenge 2004," 11 October 2006, <http://www.chatterboxchallenge.com/>.
- [182] M. Richardson and P. Domingos, "Building large Knowledge Bases by Mass Collaboration," in *K-CAP'03*, Sanibel Island, Florida, USA, 2003.
- [183] E. Mueller, "Common Sense in Humans," 20 October 2006, <http://www.signiform.com/erik/pubs/cshumans.htm>.
- [184] M. Ward, "Clicking for consciousness," BBC News, 1 August 2007, <http://news.bbc.co.uk/1/hi/sci/tech/903686.stm>.
- [185] S. Cucerzan and E. Agichtein, "Factoid Question Answering over Unstructured and Structured Web Content," Microsoft, 22 November 2006, <http://research.microsoft.com/users/silviu/Papers/trec05.pdf>.

- [186] K. Collins-Thompson, E. Terra, J. Callan, and C. Clarke, "The effect of document retrieval quality on factoid question-answering performance," in *IGIR 2004, Sheffield, UK*, 2004.
- [187] E. W. D. Whittaker, J. Hamonic, D. Yang, T. Klingberg, and S. Furui, "Monolingual Web-based Factoid Question Answering in Chinese, Swedish, English and Japanese," in *EACL 2006 Workshop on Multilingual Question Answering - MLQA06*, 2006.
- [188] E. Agichtein, S. Cucerzan, and E. Brill, "Analysis of factoid questions for effective relation extraction," in *28th annual international ACM SIGIR conference on Research and development in information retrieval*, Salvador, Brazil, 2005.
- [189] A. Franz and T. Brants, "All our N-gram are Belong to You," Google Machine Translation Team, 20 October 2006, <http://googleresearch.blogspot.com/2006/08/all-our-n-gram-are-belong-to-you.html>.
- [190] O. S. Goh and C. C. Fung, "Building an Intelligent Conversation Agent's Domain Knowledge based on a Web Knowledge Trust Model (WKTm)," *Special Issue on Knowledge Discovery for Web Intelligence, ACM Transactions on Knowledge Discovery from Data*, Submitted for review.
- [191] O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "Multilevel Natural Language Query Approach for Conversational Agent System," *International Journal of Computer Science*, vol. 33, pp. 7-13, 2007.
- [192] O. S. Goh, A. Depickere, C. C. Fung, and K. W. Wong, "Top-down Natural Language Query Approach for Embodied Conversational Agent," in *International MultiConference of Engineers and Computer Scientists 2006*, Hong Kong, 2006.
- [193] J. A. Fodor, *Elm and the Expert: An Introduction to Mentalese and Its Semantics*: Cambridge University Press, 1994.
- [194] K. R. McKeown, *Text Generation*: Cambridge University Press, 1995.
- [195] J. D. Moore and C. L. Paris, "Planning text for advisory dialogues: Capturing intentional and rhetorical information," *Computational Linguistics*, vol. 19, pp. 651-695, 1992.
- [196] D. Radev and K. McKeown, "Generating natural language summaries from multiple on-line sources," *Computational Linguistics*, vol. 24, pp. 469-500, 1998.
- [197] E. Reiter and R. Dale, *Building Natural Language Generation Systems*: Cambridge Univ. Press, 2000.
- [198] R. A. Brooks, "The Cog Project: Building a Humanoid Robot," in *1st International Conference on Humanoid Robots and Human friendly Robots*, Tsukuba, Japan, 1998.
- [199] M. A. Bickel, "Automatic correction to misspelled names: a fourth generation language approach," *CACM*, vol. 30, 1987.
- [200] J. McCarthy, "An Example for Natural Language Understanding and the AI Problems it Raises," Stanford University, 25 July 2007, <http://www-formal.stanford.edu/jmc/mrhug/mrhug.html>.
- [201] H. L. Dreyfus, *What computers still can't do – a critique of artificial reason*: MIT Press, 1992.
- [202] A. Newell, J. C. Shaw, and H. A. Simon, "Report on a general problem-solving program," in *International Conference on Information Processing*, Paris: UNESCO House, 1959, pp. 256-264.
- [203] N. I. Badler, M. S. Palmer, and B. R., "Animation Control for Real-Time Visual Humans," *Communication of the ACM*, pp. 65-73, 1999.
- [204] Y. Shinyama, T. Tokunaga, and H. Tanaka, "A Software Robot Kairai that Understands Natural Language," *Journal of JIPS*, vol. 42, pp. 1359-1367, 2001.
- [205] S. M. A. Masum and M. Ishizuka, "Integrating Natural Language Understanding and a Cognitive Approach to Textual Emotion Recognition for Generating Human-Like Responses," in *International Conference on Computer and Information Technology (ICCIT2006)*, Dhaka, Bangladesh, 2006.

- [206] K. Funakoshi, T. Tokunaga, and H. Tanaka, "Conversational Animated Agent System K3," in *International Conference on Intelligent User Interfaces (IUI2006)*, 2006.
- [207] M. Mateas and A. Stern, "Natural Language Understanding in Façade: Surface-text Processing," in *Technologies for Interactive Digital Storytelling and Entertainment (TIDSE)*, Darmstadt, Germany, 2004.
- [208] M. Mateas and A. Stern, "A behavior language for story-based believable agents," *IEEE Intelligent Systems*, vol. 17, pp. 39-47, 2002.
- [209] M. Mateas and A. Stern, "Integrating plot, character and natural language processing in the interactive drama Façade," in *International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE '03)*, Darmstadt, Germany, 2003.
- [210] O. S. Goh, C. Ardil, W. Wong, and C. C. Fung, "A Black-box Approach for Response Quality Evaluation Conversational Agent System," *International Journal of Computational Intelligence*, vol. 3, pp. 195-203, 2006.
- [211] D. Lin, "Dependency-Based Evaluation of MINIPAR," in *1st International Conference on Language Resources and Evaluation*, 1998.
- [212] W. Wong and O. S. Goh, "Evaluation of Response Quality for Heterogeneous Question Answering," in *IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005)*, Compiègne, France, 2005.
- [213] B. A. Shawar, E. Atwell, and A. Roberts, "FAQChat as an Information Retrieval System," in *2nd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznan, Poland, 2005.
- [214] R. S. Wallace, "AIML Implementations," ALICE Foundation, 25 July 2007, <http://www.alicebot.org/downloads/programs.html>.
- [215] R. S. Wallace, "AIML Overview. Online documentation," Pandorabots, 1 August 2007, <http://www.pandorabots.com/pandora/pics/wallaceaimltutorial.html>.
- [216] C. Bialik, "Just How Deadly Is Bird Flu? It Depends on Whom You Ask," *The Wall Street Journal Online*, 20 March 2006, [http://online.wsj.com/public/article/SB110512998255120225.html?mod=todays\\_free\\_feature](http://online.wsj.com/public/article/SB110512998255120225.html?mod=todays_free_feature).
- [217] S. Lovgren, "Is Asian Bird Flu the Next Pandemic?," *National Geographic News*, 20 March 2006, [http://news.nationalgeographic.com/news/2004/12/1207\\_041207\\_birdflu.html](http://news.nationalgeographic.com/news/2004/12/1207_041207_birdflu.html).
- [218] O. S. Goh, C. C. Fung, and S. H. Lau, "Object-Based Learning with Concept Mapping Embodied Intelligent Agent," *International Journal of Computer, The Internet and Management*, vol. 12, pp. 30.1 - 30.7, 2005.
- [219] O. S. Goh and C. C. Fung, "AINI - Embodied Conversation Agent Applicable for Interactive Games," in *7th WSEAS International Conference on Applied Computer and Applied Computational Science (ACACOS '08)*, Hangzhou China, 2008.
- [220] J. Jia, "The study of the application of a keywords-based chatbot system on the teaching of foreign languages," University of Augsburg 2003.
- [221] O. S. Goh, C. C. Fung, K. W. Wong, A. Depickere, and W. Wilson, "Domain Knowledge Model for Embodied Conversation Agent," in *3rd International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS 2005)*, Singapore, 2005.
- [222] O. S. Goh and C. C. Fung, "Acquiring Trustworthy Knowledge for Conversation Agents based on a Web Knowledge Trust Model," in *IAENG International Conference on Internet Computing and Web Services*, Hong Kong, 2008.
- [223] O. S. Goh, C. C. Fung, and K. W. Wong, "Query Based Intelligent Web Interaction with Real World Knowledge," *New Generation Computing*, vol. 26, pp. 3-22, 2008.

- [224] A. Franz and T. Brants, "All our N-gram are Belong to You," Google Machine Translation Team, 20 October 2006, <http://googleresearch.blogspot.com/2006/08/all-our-n-gram-are-belong-to-you.html>.
- [225] A. Kilgarriff and G. Grefenstette., "Introduction to the special issue on the web as corpus," *Computational Linguistics*, vol. 29, pp. 333–347, 2003.
- [226] L. Rainie and S. Fox, "The online health care revolution: How the Web helps Americans take better care of themselves," Pew Research Center, 9 July 2007, [http://www.pewInternet.org/pdfs/PIP\\_Health\\_Report.pdf](http://www.pewInternet.org/pdfs/PIP_Health_Report.pdf).
- [227] D. R. Gifford and D. L. Carcieri, "The Rhode Island HEALTH Web Query System," Rhode Island Department of Health, 7 July 2007, <http://www.health.state.ri.us/data/webquery.php>.
- [228] G. Eysenbach, J. Powell, O. Kuss, and E.-R. Sa, "Empirical studies assessing the quality of health information for consumers on the world wide web, a systematic review," *Journal of the American Medical Association*, vol. 287, pp. 2691-2700, 2002.
- [229] S. Fox and L. Rainie, "Vital Decisions: How Internet users decide what information to trust when they or their loved ones are sick," Pew Internet and American Life Project, Washington D.C. 2002.
- [230] E. Sillence, P. Briggs, L. Fishwick, and P. Harris, "Trust and Mistrust of Online Health Sites," in *Computer Human Interaction*, Vienna, Austria, 2004, pp. 663-670.
- [231] D. McAllister, "Affect and cognition-based trust as foundations for interpersonal cooperation in organizations.," *Academy of Management Journal*, vol. 38, pp. 24–59., 1995.
- [232] A. Parkhe, "Building trust in international alliances," *Journal of World Business*, vol. 33, pp. 417–437, 1998.
- [233] K. de Ruyter, M. Wetzels, and M. Kleijnen, "Customer adoption of e-services: an experimental study," *International Journal of Service Industry Management*, vol. 12, pp. 184–207, 2001.
- [234] P. Kanawattanachai and Y. Yoo, "Dynamic nature of trust in virtual teams," *Journal of Strategic Information Systems*, vol. 11, pp. 187–213, 2002.
- [235] K. K. Kim and B. Prabhakar, "Initialtrust and the adoption of B2C e-commerce: the case of Internet banking," in *The DATA BASE for Advances in Information Systems*, 2002.
- [236] D. Gefen, E. Karahanna, and D. W. Straub, "Trust and TAM in online shopping: an integrated model," *MIS*, vol. Quarterly 27, pp. 51–90, 2003.
- [237] S. Grabner-Kräuter and E. A. Kaluscha, "Empiricalresearch in on-line trust: a review and critical assessment," *International Journal Human-Computer Studies*, vol. 58, pp. 783–812, 2003.
- [238] D. L. Paul and R. R. McDaniel Jr., "A field study of the effect of interpersonal trust on virtual collaborative performance," *MIS Quarterly*, vol. 28 (2), pp. 183–227, 2004.
- [239] J. M. Leimeister, W. Ebner, and H. Krcmar, "Design, implementation, and evaluation of trust-supporting components in virtual communities for patients," *Journal of Management Information Systems*, vol. 21, pp. 101–135, 2005.
- [240] P. Ratnasingam, "Trust in inter-organizational exchanges: a case study in business to business electronic commerce," *Decision Support Systems*, vol. 39, pp. 525–544, 2005.
- [241] M.-H. Hsu, T. L. Jub, C.-H. Yenc, and C.-M. Chang, "Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectations," *International Journal Human-Computer Studies*, vol. 65, pp. 53–169, 2007.
- [242] M. Koufaris and W. Hampton-Sosa, "Customer trust online: examining the role of the experience with the Web-site," Baruch College, New York, 27 July 2007, <http://cisnet.baruch.cuny.edu/papers/cis200205.pdf>.



- [243] B. J. Fogg, L. Marable, J. Stanford, and E. R. Tauber, "How Do People Evaluate a Web Site's Credibility? Results from a Large Study," Consumer WebWatch & Stanford University, Yonkers, N.Y 2002.
- [244] C. L. Corritorea, B. Krachera, and S. Wiedenbeckb, "On-line trust: concepts, evolving themes, a model," *International Journal Human-Computer Studies*, vol. 58, pp. 737–758, 2003.
- [245] M. Singer, "Improving Web Trust," Internetnews, 5 July 2007, <http://www.Internetnews.com/ec-news/article.php/3408571>.
- [246] MadTek, "Evaluating Health Related Websites," MadTek, 19 april 2006, <http://www.madtek.com/content/index.pl?id=2837&isa=Category>.
- [247] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen, "Combating web spam with TrustRank," in *Thirtieth International Conference on Very Large Data Bases*, Toronto, Canada, 2004.
- [248] B. J. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, and M. Treinen, "What Makes Web Sites Credible? A Report on a Large Quantitative Study," in *ACM CHI 2001 Conference on Human Factors in Computing Systems*, Seattle, WA, USA., 2001, pp. 61-68.
- [249] R. Dhamija, J. D. Tygar, and M. Hearst, "Why Phishing Works," in *CHI 2006*, Montréal, Québec, Canada., 2006.
- [250] WebWatch, "A Matter of Trust: What Users Want From Web Sites," Consumer Reports WebWatch, 1 May 2007, <http://www.consumerwebwatch.org/pdfs/a-matter-of-trust.pdf>.
- [251] R. T. Stephens, "A Framework for the Identification of Electronic Commerce Design Elements that Enable Trust," Nova Southeastern University, 7 July 2007, <http://www.rtodd.com/speaking/2003/ORMS2003/Research%20Study%20v1.pdf>.
- [252] B. J. Fogg, T. Kameda, J. Boyd, J. Marshall, R. Sethi, M. Sockol, and T. Trowbridge, "Stanford-Makovsky Web Credibility Study 2002: Investigating what makes Web sites credible today," A Research Report by the Stanford Persuasive Technology Lab in collaboration with Makvosky & Company. Stanford University. 2002.
- [253] B. J. Fogg and H. Tseng, "The Elements of Computer Credibility," in *CHI99 Conference on Human Factors and Computing Systems*, 1999, pp. 80-87.
- [254] P. A. Pavlou, "Integrating trust in electronic commerce with the technology acceptance model: model development and validation," in *Seventh Americas Conference on Information Systems*, Atlanta, GA, 2001, pp. 816–822.
- [255] V. Shankar, G. L. Urban, and F. Sultan, "Online trust: a stakeholder perspective, concepts, implications, and future directions," *Journal of Strategic Information Systems*, vol. 11, pp. 325–344, 2002.
- [256] N. Bos, J. Olson, D. Gergle, G. Olson, and Z. Wright, "Effects of Four Computer-Mediated Communications Channels on Trust Development," in *CHI 2002*, Minneapolis, Minnesota, USA, 2002.
- [257] M. J. Ashleigh and J. Nandhakumar, "Trust and technologies: Implications for organizational work practices," *Decision Support Systems*, vol. 43, 2007.
- [258] H. C. Wong and K. Sycara, "Adding security and trust to multi-agent systems," in *Autonomous Agents '99: Workshop on Deception, Fraud and Trust in Agent Societies*, Seattle, WA, 1999, pp. 149–161.
- [259] J. Cassell and T. Bickmore, "External Manifestations of Trustworthiness in the Interface," *Communication of the ACM*, vol. 43, 2000.
- [260] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, pp. 319-340, 1989.
- [261] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks and ISDN Systems*, vol. 30, pp. 107–117, 1998.

- [262] F. Keller, M. Lapata, and O. Ourioupina, "Using the Web to Overcome Data Sparseness," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Philadelphia, 2002, pp. 230-237.
- [263] A. Bifet and C. Castillo, "An Analysis of Factors Used in Search Engine Ranking," in *14th International World Wide Web Conference*, Chiba, Japan, 2005, pp. 70-78.
- [264] S. Lawrence and C. L. Giles, "Searching the World Wide Web," *Science*, vol. 280, pp. 98-100, 1998.
- [265] S. Lawrence and C. L. Giles, "Accessibility of information on the web," *Nature*, vol. 400, pp. 107-109, 1999.
- [266] M. L. Baker, "Open-Source Initiative Targets Bird Flu," Ziff Davis Publishing Holdings Inc, 29 April 2007, <http://www.eweek.com/article2/0,1895,1963157,00.asp>.
- [267] A. Z. Broder, M. Najork, and J. L. Wiener, "Efficient URL Caching for World Wide Web Crawling," in *WWW 2003*, Budapest, Hungary., 2003.
- [268] M. Baroni and M. Ueyama, "BootCaT: Bootstrapping corpora and terms from the web," in *Fourth Language Resources and Evaluation Conference*, 2004.
- [269] S. Tseng and B. J. Fogg, "Credibility and Computing Technology," *Communications of the ACM*, vol. 42, pp. 39-44, 1999.
- [270] N. Kim and B. J. Fogg, "World Wide Web Credibility: What effects do advertisements and typos have on the perceived credibility of Web page information?," Stanford University, 1999.
- [271] H. Finberg, H. Stone, and D. Lynch, "Digital Journalism Credibility Study," 28 April 2007, [http://www.journalists.org/Programs/credibility\\_study.pdf](http://www.journalists.org/Programs/credibility_study.pdf).
- [272] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, vol. 9, pp. 60-62, March 1994 1994.
- [273] T. Dunning, "Accurate methods for the statistics of surprise and coincidence," *Computational Linguistics*, vol. 19, pp. 61-74, 1993.
- [274] S. L. Piao, G. Sun, P. Rayson, and Q. Yuan, "Automatic extraction of Chinese multiword expressions with a statistical tool," in *Workshop on Multi-word-expressions in a Multilingual Context held in conjunction with the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, Trento, Italy, 2006, pp. 17-24.
- [275] W. Wong, O. S. Goh, M. I. Desa, and S. Sahib, "Online Cyberlaw Knowledge Base Construction Using Semantic Network," in *International Conference on Computational Intelligence for Modelling, Control and Automation*, Rhodes, Greece, 2004, pp. 347-352.
- [276] S. Sharoff, "Creating General-Purpose Corpora Using Automated Search Engine Queries," in *WaCky! Working papers on the Web as Corpus*, M. Baroni and S. Bernardini, Eds., 2006.
- [277] D. Sleator and D. Temperley, "Parsing English with a Link Grammar," Carnegie Mellon University Computer Science Technical report CMU-CS-91-196, 1991.
- [278] D. Klein and C. Manning, "Accurate Unlexicalized Parsing," in *41st Annual Meeting of the Association for Computational Linguistics*, 2003, pp. 423-430.
- [279] S. Bangalore, C. Doran, B. Hockey, and A. Sarkar, "Grammar and Parser Evaluation in the XTAG Project," in *1st International Conference on Language Resources and Evaluation*, Granada, Spain, 1998.
- [280] ALICE, "ALICE Loebner Prize Winner Transcripts," Loebner Prize Homepage, 20 August 2007, [http://loebner.net/Prize/2001\\_Contest/Richard\\_Wallace.txt](http://loebner.net/Prize/2001_Contest/Richard_Wallace.txt).
- [281] S. E. Journal, "Google and Yahoo Search Engine Technology Comparison," Search Engine Journal, 3 August 2006, <http://www.searchenginejournal.com/?p=2267>.
- [282] M. Bianchini, M. Gori, and F. Scarselli, "PageRank and web communities," in *Web Intelligence Conference*, 2003.

- [283] Google, "Google SOAP Search API (beta)," Google, 3 August 2006, <http://www.google.com/apis/>.
- [284] Yahoo!, "Yahoo! Search Web Services," Yahoo!, 3 August 2006, <http://developer.yahoo.com/search/>.
- [285] J. Lin and B. Katz, "Building a Reusable Test Collection for Question Answering," *Journal of the American Society of Information Science and Technology*, vol. 57, pp. 851-861, 2006.
- [286] W. C. Ogden and P. Bernick, "Using Natural Language Interfaces," in *Handbook of Human-Computer Interaction*, M. Helander, Ed. North-Holland: Elsevier, 1996.
- [287] S. Livingstone, "UK Children Go Online: Surveying the experiences of young people and their parents," 1 November 2006, [www.lse.ac.uk/collections/children-go-online/UKCGO\\_Final\\_report.pdf](http://www.lse.ac.uk/collections/children-go-online/UKCGO_Final_report.pdf).
- [288] C. Neustaedter, "A 3D Instant Messenger Visualization Using a Space Metaphor," 1 November 2006, [http://pages.cpsc.ucalgary.ca/~carman/GLabIMVis/IMVis\\_Paper.pdf](http://pages.cpsc.ucalgary.ca/~carman/GLabIMVis/IMVis_Paper.pdf).
- [289] A. F. Rovers and H. A. Van Essen, "Him: A framework for haptic instant messaging," in *CHI*, Viena, April 2004.
- [290] S. M. Cherry, "IM Means Business," in *IEEE Spectrum Magazine*, November 2002.
- [291] J. D. Herbsleb, D. L. Atkins, D. G. Boyer, M. Handel, and T. A. Finholt, "Introducing Instant Messaging and Chat in the Workplace," in *CHI'2002*, Minneapolis, Minnesota, USA., 2002.
- [292] J. O'Neill and D. Martin, "Text Chat In Action," in *GROUP'03*, Sanibel Island, Florida, USA., 2003.
- [293] B. Nardi, S. Whittaker, and E. Bradner, "Interaction and Outeraction: Instant Messaging in Action," in *CSCW'2000*, 2000, pp. 79-88.
- [294] R. C. Forgas and J. S. Negre, "The use of new Technologies amongst minors in the Balearic Islands.," in *IAARE Conference*, Melbourne, 2004.
- [295] D. Craig, "Instant messaging: the language of youth literacy. The Boothe Prize Essays," 1 November 2006, [www.stanford.edu/group/pwr/publications/Boothe\\_0203/PWR%20Boothe-Craig.pdf](http://www.stanford.edu/group/pwr/publications/Boothe_0203/PWR%20Boothe-Craig.pdf).
- [296] Y. Hård af Segerstad and S. Sofkova Hashemi, "Exploring the Writing of Children and Adolescents in the Information Society EARLI SIG Writing," in *9th International Conference of the EARLI - Special Interest Group on Writing*, Geneve, Switzerland, 2004, pp. 20-22.
- [297] S. R. Crenzel and V. L. Nojima, "Children and instant messaging," in *IEA 2006: 16th World Congress on ergonomics*, Maastricht, Netherlands, 2006.
- [298] C. Saunders, "Vendors Debut New IM Bot Tech," 1 December 2006, [http://www.instantmessagingplanet.com/enterprise/article.php/11208\\_1575161](http://www.instantmessagingplanet.com/enterprise/article.php/11208_1575161).
- [299] A. R. Hickey, "IM bots ease access to corporate apps," 1 December 2006, [http://searchmobilecomputing.techtarget.com/originalContent/0,289142,sid40\\_gci1136486,00.html?bucket=NEWS&topic=299728](http://searchmobilecomputing.techtarget.com/originalContent/0,289142,sid40_gci1136486,00.html?bucket=NEWS&topic=299728).
- [300] T. Gray, "The Brokerage and The Bot," 1 December 2006, <http://www.ecrmguide.com/article.php/3494791>.
- [301] D. Chin, "An Analysis of Scripts Generated in Writing Between Users and Computer Consultants," in *National Computer Conference*, 1984, pp. 637-642.
- [302] R. Guindon, "How to interface to advisory systems? Users request help with a very simple language," in *SIGCHI conference on Human factors in computing systems*, Washington, D.C, 1988.
- [303] D. N. Chin, "Empirical Evaluation of User Models and User-Adapted Systems," *User Modeling and User-Adapted Interaction*, vol. 11, pp. 181-194, 2001.
- [304] S. Whittaker and P. Stenton, "User studies and the design of natural language systems," in *Fourth conference on European chapter of the Association for Computational Linguistics*, Manchester, England, 1989.

- [305] E. Isaacs, A. Walendowski, S. Whittaker, D. J. Schiano, and C. Kamm, "The Character, Functions, and Styles of Instant Messaging in the Workplace," in *CSCW '02*, NY, 2002, pp. 11-20.
- [306] D. Avrahami and S. E. Hudson, "Communication Characteristics of Instant Messaging: Effects and Predictions of Interpersonal Relationships," in *CSCW'06*, Banff, Alberta, Canada, 2006.
- [307] A. Lüdeling, S. Evert, and M. Baroni, "Using Web data for linguistic purposes," in *Corpus linguistics and the Web.*, M. Hundt, N. Nesselhauf, and C. Biewer, Eds. Amsterdam: Rodopi, 2007, pp. 7-24.
- [308] O. S. Goh, A. Depickere, C. C. Fung, and K. W. Wong, "An Analysis of Corpus from Human Computer Exchanges Using MSN Messenger," in *IAENG International Conference on Internet Computing and Web Services (ICICWS'07)*, Hong Kong, 2007.
- [309] N. K. Anagnostou and G. R. S. Weir, "Review of software applications for deriving collocations," in *ICT in the Analysis, Teaching and Learning of Languages*, Glasgow, UK, 2006.
- [310] L. F. Costa, "Esfinge - A Question Answering System in the Web using the Web," in *Demonstration Session of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 2006.
- [311] K. Forbes-Riley and D. J. Litman, "Using Bigrams to Identify Relationships Between Student Certainty States and Tutor Responses in a Spoken Dialogue Corpus," in *6th SIGdial Workshop on Discourse and Dialogue*, Lisbon, Portugal, 2005.
- [312] D. Inkpen and G. Hirst, "Building and Using a Lexical Knowledge Base of Near-Synonym Differences," *Computational Linguistics*, vol. 32, pp. 223-262, 2006.
- [313] T. Pedersen and S. Banerjee, "The Design, Implementation, and Use of the Ngram Statistics Package," in *Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, 2003.
- [314] J. Preiss, "Probabilistic word sense disambiguation," *Computer Speech & Language*, vol. 18, pp. 319-337, 2004.
- [315] B. Kis, B. Villada, G. Bouma, G. Ugray, T. Bíró, G. Pohl, and J. Nerbonne, "A New Approach to the Corpus-based Statistical Investigation of Hungarian Multi-Word Lexemes," in *4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal, 2004, pp. 1677-1681.
- [316] T. van der Wouden, I. Schuurman, M. Schoupe, and H. Hoekstra, "Harvesting Dutch trees: Syntactic properties of Spoken Dutch," in *Computational Linguistics in the Netherlands 2002*, Rodopi, Amsterdam, 2003, pp. 129-141.
- [317] S. Koeva, "Multi-word Term Extraction for Bulgarian," in *Workshop on Balto-Slavonic Natural Language Processing*, Prague, Czech Republic, 2007, pp. 59-66.
- [318] S. Banerjee and T. Pedersen, "The Design, Implementation, and Use of the Ngram Statistic Package," in *Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, 2003.
- [319] J. Donath, "A semantic approach to visualizing online conversations," *Communications of the ACM*, vol. 45, pp. 45-49, 2002.
- [320] S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*. San Francisco, USA: Morgan Kaufmann, 1999.
- [321] P. Neumann, S. Schlechtweg, and S. Carpendale, "Arctrees: Visualizing relation in hierarchical data," in *Eurographics - IEEE VGTC Symposium on Visualization*, 2005.
- [322] N. Rojas, "MSN history visualization," Rhizome.org, 12 August 2007, <http://rhizome.org/object.php?45393>.
- [323] O. S. Goh, C. C. Fung, and K. W. Wong, "VisualChat: A Visualisation Tool for Human-Machine Interaction," in *2008 IEEE/WIC/ACM Intelligent Web Interaction Workshops (IWI'08)*, Sydney, Australia, Submitted for review.



- [324] A. Garcia and J. Jacobs, "The eyes of the beholder : Understanding the turn-taking system in quasi-synchronous computer mediated communication.," *Research on Language and Social Interaction*, vol. 32, pp. 337-367, 1999.
- [325] C. Alphonso, "Texting helps teens' grammar," Bell Globemedia, 1 November 2006, <http://www.theglobeandmail.com/servlet/story/LAC.20060801.TEXT01/TPStory/TPNational/Ontario/>.
- [326] S. L'Abbé, "Instant msg-ing messes with grammar? As if! lol! Teens adopting unique linguistic shorthand but not ruining syntax," University of Toronto, 1 November 2006, <http://www.news.utoronto.ca/bin6/060731-2474.asp>.
- [327] J. Stewart, "Word Frequency usage in English," 5 November 2006, <http://www.write101.com/W.Tips74.htm>.
- [328] C. Nass, "Etiquette inequality: Exhibitions and expectations of computer politeness.," *Communications of the ACM*, vol. 47, pp. 35-37, 2004.
- [329] G. Leech, P. Rayson, and A. Wilson, *Word Frequencies in Written and Spoken English: based on the British National Corpus*. London: Longman, 2001.
- [330] R. Gunning, "Plain Language At Work Newsletter," 23 March 2004, <http://www.impact-information.com/impactinfo/newsletter/plwork08.htm>.
- [331] R. Gunning, *The Technique of Clear Writing*: McGraw-Hill, 1952.
- [332] O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "Gunning-Fog Index for Instant Messaging in the Embodied Conversational Agent," in *The 3rd International Conference on Natural Computation (ICNC'07)*, Haikou, China, 2007.
- [333] A. T. Rabin, "Determining difficulty levels of text written in languages other than English," in *Readability: Its past, present, and future*, L. Zakaluk and S. J. Samuels, Eds. Newark, DE: International Reading Association, 1988.
- [334] A. L. Sherman, *Analytics of literature: A manual for the objective study of English prose and poetry*. Boston: Ginn & Co, 1893.
- [335] W. H. DuBay, "The Principles of Readability," Costa Mesa, CA: Impact Information, 10 November 2006, <http://www.impact-information.com/impactinfo/readability02.pdf>.
- [336] J. M. Spool, T. Scanlon, W. Schroeder, C. Snyder, and T. DeAngelo, *Web Site Usability: A Designer's Guide*: Morgan Kaufmann Publishers, Inc., San Francisco, 1999.
- [337] T. Sikorski and a. Allen, "A task-based evaluation of the TRAINS-95 dialogue system," in *ECAI Workshop on Dialogue Processing in Spoken Language Systems.*, 1996.
- [338] K. Isbister, H. Nakanishi, and T. Ishida, "Helper Agent: Designing and Assistant for Human-Human Interaction in a Virtual Meeting Space," in *Human Factors in Computing Systems (CHI2000)*, New York, 2000, pp. 57-64.
- [339] T. Koda and T. Ishida, "Cross-cultural Comparison of Interpretation of Avatars' Facial Expressions.6," in *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, 2006, pp. 130-136.
- [340] O. S. Goh, C. C. Fung, and K. W. Wong, "Paralinguistic Cues in IM Conversation Robots," *International Journal of Social Robotics*, Submitted for review
- [341] M. Pleari and C. L. Lisetti, "Toward Multimodal Fusion of Affective Cues," in *HCM'06*, Santa Barbara, California, USA, 2006.
- [342] M. Argyle and M. G. Cook, *Mutual Gaze*. Cambridge: Cambridge University Press, 1976.
- [343] E. Isaacs and J. Tang, "What Video Can and Can't Do for Collaboration: A Case Study," in *Multimedia Systems*, 1994.
- [344] N. S. Baron, "Instant Messaging by American College Students A Case Study in Computer-Mediated Communication," American Association for the Advancement of Science, 10 August 2007, <http://www.american.edu/tesol/Baron-AAAS-IM%20by%20American%20College%20Students.pdf>.


- [345] V. Bruce and A. Young, *In the eye of the beholder: The science of face perception.*: Oxford University Press, UK, 1998.
- [346] P. Ekman and W. Friesen, "The repertoire of nonverbal behavior. Categories, origins, usage, and coding.," *Semiotica*, vol. 1, pp. 49-98, 1969.
- [347] F. Viégas, "Visualization Email Content: Portraying Relationships from Conversational Histories," in *CHI2006*, 2006.
- [348] O. S. Goh, C. C. Fung, K. W. Wong, and A. Depickere, "Linguistic Analysis of an Instant Messaging Conversation Agent Corpus," *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, Submitted for review.
- [349] C. T. Ishi, H. Ishiguro, and N. Hagita, "Using Prosodic and Voice Quality Features for Paralinguistic Information Extraction," 2006.
- [350] D. C. O'Connell and S. Kowal, "Uh and Um Revisited: Are They Interjections for Signaling Delay?," *Journal of Psycholinguistic Research*, vol. 34, pp. 555-576, 2005.
- [351] J. Hallebeek, *A Formal Approach to Spanish Syntax*. Netherlands: Rodopi, 1992.
- [352] M. Federico, "A system for the retrieval of Italian broadcast news," *Speech Communication*, vol. 32, pp. 37-47 2000.
- [353] B. Boneva, A. Quinn, R. Kraut, S. Kiesler, J. Cummings, and I. Shklovski, "Teenage Communication in the Instant Messaging Era," in *Information Technology at Home*, R. Kraut, M. Brynin, and S. Kiesler, Eds. New York:: Oxford University Press, 2006, pp. 612-72.
- [354] D. Crystal, *Language and the Internet*: Cambridge University Press, 2001.
- [355] E. Jansen, "Netlingo: The Internet Dictionary," 1 November 2006, <http://www.netlingo.com>.
- [356] T. Kontzer, "IM acronyms gain in business," ITNews, 12 November 2006, <http://www.itnews.com.au/newsstory.aspx?CIaNID=13771>.
- [357] C. Werry, "Linguistic and interactional features of Internet Relay Chat," in *Computer-mediated communication: Linguistic: social and cross-cultural perspectives*, S. C. Herring, Ed. Amsterdam: John Benjamins., 1996, pp. 47-64.
- [358] B. Levine, "IMing Is Creative Language, Study Says," Sci-Tech Today, 1 November 2006, [http://www.sci-tech-today.com/story.xhtml?story\\_id=00100040PAQP](http://www.sci-tech-today.com/story.xhtml?story_id=00100040PAQP).
- [359] S. McCloud, *Understanding Comics*: HarperPerennial, 1993.
- [360] L. Zhou and Z. Zhuang, "Recovering Punctuations in Instant Messages," in *Tenth Americas Conference on Information Systems*, New York, 2004.
- [361] L. Hamill, "Controlling Smart Devices in the Home," *The Information Society*, vol. 22, pp. 241-249, 2006.
- [362] R. Al-Halimi and R. Kazman, "Temporal indexing through lexical chaining," in *WordNet: An Electronic Lexical Database*, C. Fellbaum, Ed.: MIT Press, 1998.
- [363] Dias-da-Silva and B. Carlos, "Human language technology research and the development of the Brazilian Portuguese wordnet," in *17th International Congress of Linguists*, Prague, Czech Republic, 2003.
- [364] R. Amaro, R. P. Chaves, P. Marrafa, and S. Mendes, "Enriching Wordnets with new Relations and with Event and Argument Structures," in *Seventh International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, 2006, pp. 28-40.
- [365] R. W. Picard, *Affective Computing*. Cambridge, MA: MIT Press, 1997.
- [366] P. Salovey and J. D. Mayer, "Emotional Intelligence," *Imagination, Cognition and Personality*, vol. 9, pp. 185-211, 1990.
- [367] T. Takahashi, S. Nakanishi, Y. Kuno, and Y. Shirai, "Human-robot interface by verbal and nonverbal communication," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1998, pp. 924-929.
- [368] Y. Nakauchi, P. Naphattalung, T. Takahashi, T. Matsubara, and E. Kashiwagi, "Proposal and evaluation of natural language human-robot interface system based on

- conversation theory," in *IEEE International Conference on Robotics and Automation*, 2003, pp. 380 - 385.
- [369] O. Lemon, A. Bracy, A. Gruenstein, and S. Peters, "A multi-modal dialogue system for human-robot conversation," in *Scnd Meeting of the North American Chapter of the Association for Computational Linguistics NAACL*, 2001.
  - [370] C. L. Sidner and M. Dzikovska, "Human - Robot Interaction: Engagement between Humans and Robots for Hosting Activities," in *Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02)*, 2002.
  - [371] S. Yamada and T. Yamaguchi, "Training AIBO like a Dog," in *13th International Workshop on Robot and Human Interactive Communication (ROMAN-2004)*, Kurashiki, Japan, 2004.
  - [372] J. D. Pettit and T. Bräunl, "A Framework for Cognitive Agents," *Transactions on Control, Automation, and Systems Engineering, ICASE*, vol. 2, pp. 1-8, 2003.
  - [373] J. Bos, E. Klein, and T. Oka, "Meaningful Conversation with a Mobile Robot," in *10th European Chapter of the Association for Computational Linguistics (EACL)*, Budapest, Hungary, 2003.
  - [374] T. Bräunl, *Mobile Robot Design and Applications with Embedded Systems*, 2nd ed. Berlin, Heidelberg: Springer-Verlag, 2006.
  - [375] G. Hoffman and C. Breazeal, "Effects of Anticipatory Action on Human-Robot Teamwork Efficiency, Fluency and Perception of Team," in *ACM/IEEE International conference on Human-Robot Interaction (HRI'07)*, Arlington, Virginia, USA, 2007.


## APPENDIXES

### APPENDIX A: Online Consent Form Approved by Human Research Ethics Committee (HREC), Murdoch University


**ANNOUNCEMENT**



Read the Announcement carefully



Click "NEXT" button to continue



Click "CLOSE" button to stop the study

Artificial Intelligent Neural-network Identity - Mozilla Firefox

File Edit View Go Bookmarks Yahoo! Tools Help

http://ainibot.murdoch.edu.au/study/

Getting Started Latest Headlines School of Law WebTV Law.TV

Search Web Upgrade to the latest Yahoo! Toolbar Mail My Yahoo!

**CAs**  
**Conversation Agents**  
Artificial Intelligence Natural-language Identity

**ANNOUNCEMENT**

Calling all Instant Messaging users and chatterbots developers. You can help in this study by consenting to participate in a research on embodied conversational agents software robot or chatterbots development.

To participate, read the Consent Statement and follow the instructions. This study will lead you to an online chat session with chatterbots using MSN Messenger. Identities of all participants will remain confidential. No name-identified records will be recorded in this study.

NEXT CLOSE


©2005-2008 http://ainibot.murdoch.edu.au All rights reserved..

Welcome to Artificial Intelligent Neural-Netw

Microsoft Internet Explorer

The Web page you are viewing is trying to close the window.  
Do you want to close this window?

Yes No



## CONSENT STATEMENT



Read and understand the "Consent Statement" carefully



Click "NEXT" button to continue



Click "CLOSE" button to stop the study

Artificial Intelligent Neural-network Identity - Mozilla Firefox

File Edit View Go Bookmarks Yahoo! Tools Help

http://aini.murdoch.edu.au/study/consent\_statement.php

Getting Started Latest Headlines School of Law WebTV Law.TV

Search Web Upgrade to the latest Yahoo! Toolbar Mail My Yahoo!

msn\_history : Built with Processing Artificial Intelligent Neural-network I...

---

CAs  
Artificial Intelligent Neural-network Identity

### CONSENT STATEMENT

My name is Ong Sing Goh. I am a Ph.D. candidate at the School of Information Technology, Murdoch University, Western Australia as part of my research; I am conducting a study regarding the embodied conversational agent or chatterbots called Artificial Intelligent Natural-language Identity (AINI).

Your involvement in this project involves discussing about linguistics and communication features between online users and AINI through MSN Messenger. The time required will all depend on how active you are in posting and reading messages. Participants can decide to withdraw their consent at any time.

Please be advised that your participation is voluntary and that you may withdraw at any time or decline to answer any question posted by the AINI. We would also like to assure you that your responses will be held strictly confidential. All information given during the study is confidential and no names or other information that might identify you will be used in any publication arising from the research. Feedback on the study will be provided to participants. Only persons over 18 years of age may participate in this study.

By accept an invitation as a AINI's buddies, you are stating that you meet the age requirement (18 years old and above) and agree to the rules of this study. You may end your participation at any time. If you have questions, please do not hesitate to contact me Ong Sing Goh or my supervisor, Dr. Lance Fung with the following details:

Ong Sing Goh  
Room 6, E&H Transportable 1, School of Information Technology, Murdoch University  
South Street, Murdoch, Western Australia 6150, AUSTRALIA  
Phone: +61 8 9360 7507 Fax: +61 8 9360 2941  
E-mail: [o.sgoh@murdoch.edu.au](mailto:o.sgoh@murdoch.edu.au)

or

Dr Lance Chun Che Fung  
Associate Professor  
ECL 3.042, School of Information Technology, Murdoch University  
South Street, Murdoch, Western Australia 6150, AUSTRALIA  
Phone +61 8 9360 7586 Fax: +61 8 9360 2941  
Email: [L.Fung@murdoch.edu.au](mailto:L.Fung@murdoch.edu.au)

My supervisor and I are happy to discuss with you any concerns you may have on how this study has been conducted, or alternatively you can contact Murdoch University's Human Research Ethics Committee at [human.ethics@murdoch.edu.au](mailto:human.ethics@murdoch.edu.au). This consent form will be collected online to the commencement of the study.

NEXT CLOSE

©2005-2008 <http://aini.murdoch.edu.au> All rights reserved.

Welcome to Artf





## CONSENT FORM



If you have read and understand the "Consent Form", type the CONSENT KEY provided



Click "NEXT" button to continue



Click "CLOSE" button to stop the study

*The CONSENT KEY is generated randomly using CAPTCHA algorithm. CAPTCHA are used to prevent spammers or software agent from performing actions which might be used to make a profit on the part of the person running a agent.*



**Wrong CONSENT KEY.**



## CONSENT FORM



If you typed wrong  
CONSENT KEY, you  
have to type a new  
CONSENT KEY  
again.



Click "NEXT" button  
to continue



Click "CLOSE"  
button to stop the  
study

Artificial Intelligent Neural-network Identity - Mozilla Firefox

File Edit View Go Bookmarks Yahoo! Tools Help

http://ainibot.murdoch.edu.au/study/consent\_form.php

Getting Started Latest Headlines School of Law WebTV Law.TV

msn\_history : Built with Processing Artificial Intelligent Neural-network I...

**CAs**  
**Conversation Agents**  
Artificial Intelligence Natural-Language Identity

**CONSENT FORM**

I have read the information in the Consent Statement. Any questions I have asked have been answered to my satisfaction. I consent that I am above 18 years old and agree to take part in this study, however, I know that I may change my mind and stop at any time.

I understand that all information provided is treated as confidential and will not be released by the investigator unless required to do so by law.

You typed a wrong CONSENT KEY, please type again... How:

The characters are case-sensitive.

©2005-2008 http://ainibot.murdoch.edu.au All rights reserved.

Welcome to Artificial Intelligent Neural-Network Identity (

*New CONSENT KEY  
has been generated.*

Microsoft Internet Explorer

? The Web page you are viewing is trying to close the window.  
Do you want to close this window?

**Correct CONSENT KEY**



## START CHATTING



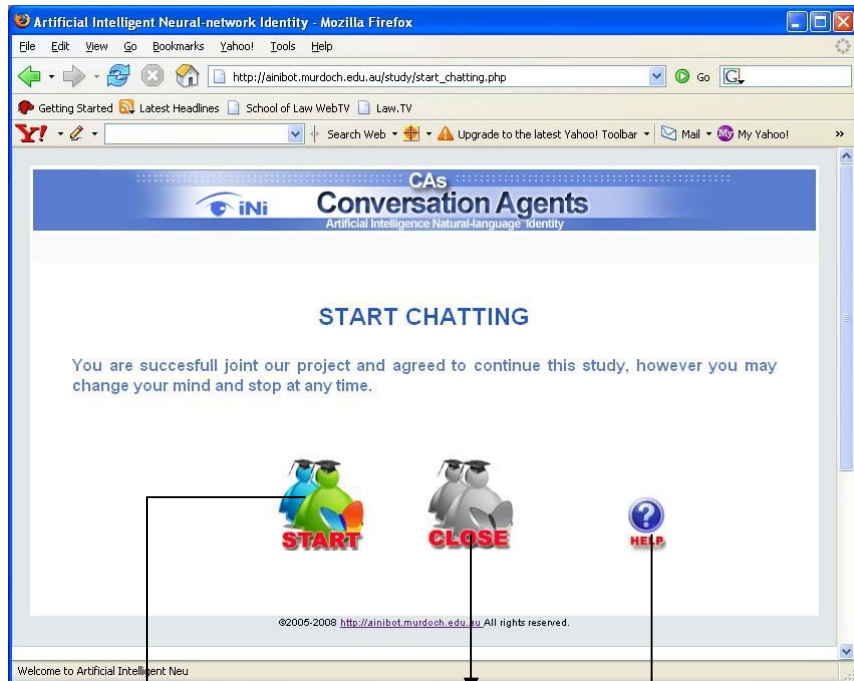
Click “START” button to start a chat session



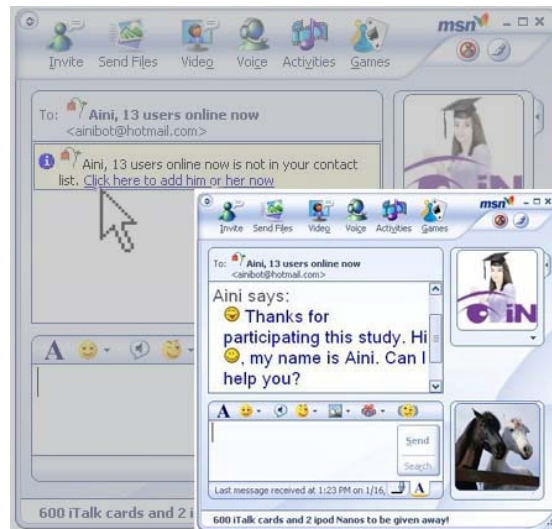
Click “CLOSE” button to stop the study



Click “HELP” button to find out how to add AINI in your MSN Messenger Contact list or check System Requirements for this study.



*When participants agreed to add AINI as his/her buddy for the first time, an alert will send to the buddy notifying them of the participation in the study. Participants are freely to chat any topics. During chatting session, participants are not restricted to any questions or regulations.*





## HELP



To add AINI in your MSN Messenger Contact list, choose "Click here to add him or her now".



You are freely to ask any questions or discuss any topics with AINI




If you are offline, please Sign-In to your messenger service first.

http://ainibot.murdoch.edu.au - Artificial Intelligent Ne...


## HELP

MSN Messenger is an instant-messaging program and you are allows to invite AINI in your list to meet over the Internet. You have full control over who you want to add to your contact list and you can easily block or remove AINI if you choose to do so. To invite AINI in your contact list for the first time, follow these two easy steps:

- 1 To add AINI in your MSN Messenger Contact List, choose "Click here to add him or her now" as below:



- 2 An alert message will send to you and notifies you that messaging session started.



**System Requirements:**

- MSN Messenger Client & Member of MSN
- If you do not have the required software, download [MSN Messenger client](#) and to be a member of MSN, [signed up](#) for its free services.

©2005-2008 <http://ainibot.murdoch.edu.au> All rights reserved.

Done

CLOSE

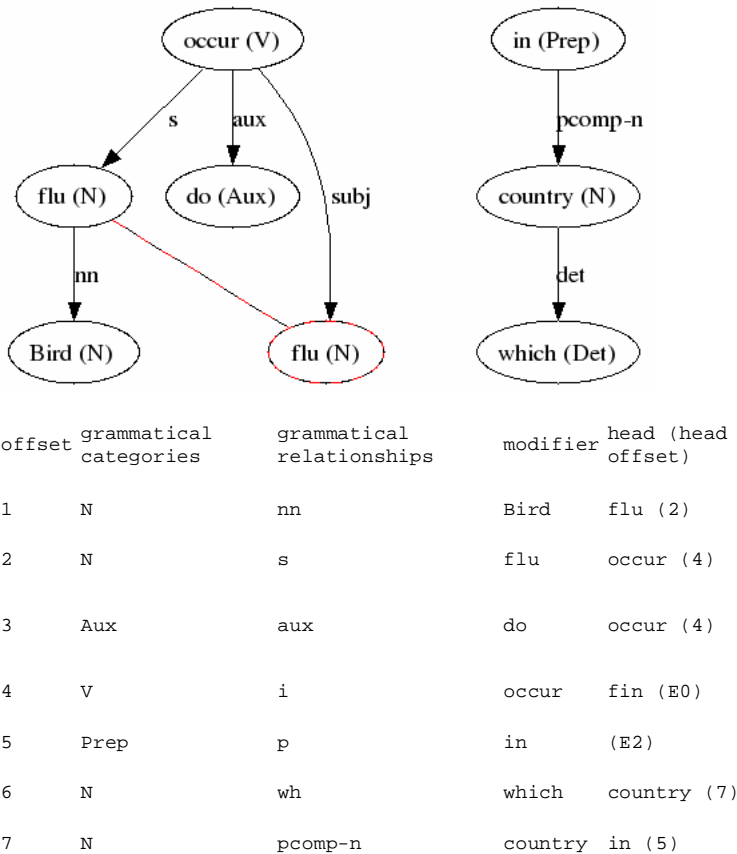
Close the  
HELP  
window.

APPENDIX B: An Example of the Full Sentence Parsing using Natural Language Understanding and Reasoning (NLUR) Module for a sentence “Bird flu did occur in which countries?”

SENTENCE PARSING  
INPUT SENTENCE 0  
Bird flu did occur in which countries?

OUTPUT OF X-MINIPAR  
(  
E2 (( ) U \* )  
E0 (( ) fin C E2 )  
1 (Bird ~ N 2 nn (gov flu))  
2 (flu ~ N 4 s (gov occur))  
3 (did do Aux 4 aux (gov occur))  
4 (occur ~ V E0 i (gov fin))  
E3 (( ) flu N 4 subj (gov occur) (antecedent 2))  
5 (in ~ Prep E2 p)  
6 (which ~ Det 7 det (gov country))  
7 (countries country N 5 pcomp-n (gov in))  
)

Visualisation of X-MINIPAR Dependency Graph



NAMED-ENTITY RECOGNITION FOR query (SENTENCE 0 )

GRAMMATICAL CATEGORIES & RELATIONSHIPS				
INPUT	1	N	nn	Bird
	2	N	s	flu
	3	Aux	aux	do
	4	V	i	occur
	5	Prep	p	in
	6	N	wh	which
	7	N	pcomp-n	country
				flu (2)
				occur (4)
				occur (4)
				fin (E0)
				(E2)
				country (7)
				in (5)

# NOUN PHRASE CHUNKING

```
<<(START-OF-HASH) << Bird(1) <<flu(2) >>do(3) >> occur(4) >> in(5) >> which(6) >> country(7)
>> (END-OF-HASH) >>
```

```
<<(START-OF-HASH) <<Bird(1) <<flu(2) <<do(3) <<occur(4) <<in(5) << which(6) <<country(7)
>>(END-OF-HASH) >>
```

**Bird(1)** is a noun but has already been use in existing NP Bird flu(1,2). **consider no more.**

**do(3)** is not a noun. **consider no more.**

**which(6)** is a noun and has not been used in existing NP Bird flu(1,2). considering for new NP.

**which(6)** is a noun but has already been use in existing NP which country(6,7). **consider no more.**

## NOUN PHRASE CHUNKING OUTPUT

2 Bird **flu**

7 which **country**

### CATEGORY ASSIGNMENT

ASSIGNMENT STARTS: Using the head **flu** to look for a match in gazetteer

#### FIRST PASS FOR **flu**

no direct match for **flu**

#### SECOND PASS is carried out

the head of noun phrase **flu** cannot be further tokenized. using the entire noun phrase **Bird flu**

using token **Bird**. try to find a partial match.

using token **flu**. try to find a partial match.

the token **flu** has a positive match from a generic (triggering word): flu.

trying to fit the token **flu** to the pattern: **(.\*\s{TOKEN})**.

instantiating the pattern to obtain **(.\*\sflu)**.

the instantiated pattern matches the noun phrase **Bird flu**

attributes identified from this instance

=> **disease\_name: Bird flu**

**flu** is tagged as **disease**.

**flu** produces the following attributes **disease\_name(X,Bird flu)**.

ASSIGNMENT ENDS

ASSIGNMENT STARTS: Using the head **country** to look for a match in gazetteer

#### FIRST PASS FOR **country**

no direct match for **country**

#### SECOND PASS is carried out

the head of noun phrase **country** cannot be further tokenized. using the entire noun phrase **which country**

using token **which**. try to find a partial match.

using token **country**. try to find a partial match.

no specific categories can be assigned to **which country**

**which country** will assigned with the default category **variable(var\_desc(X,which country))**

ASSIGNMENT ENDS

2 Bird flu	disease(disease_name(X,Bird flu))	flu
7 which country	variable(var_desc(X,which country))	country

### RELATION INFERENCE FOR query (SENTENCE 0)

#### INPUT GRAMMATICAL CATEGORIES & RELATIONSHIPS

1 N	nn	Bird	flu (2)
2 N	s	flu	occur (4)
3 Aux	aux	do	occur (4)
4 V	i	occur	fin (E0)
5 Prep	p	in	(E2)
6 N	wh	which	country (7)

```

7 N                                pcomp-n                                country in (5)
NOUN PHRASES & THEIR CATEGORIES
2 Bird flu                        disease(disease_name(X,Bird flu))        flu
7 which country                  variable(var_desc(X,which country))      country

TYPE 1 (POSSESSION) RELATIONS
finding for tokens that have gen relation modifying a head

TYPE 2 (APPOSITIVE) RELATIONS
finding for tokens that have appo relation modifying a head

TYPE 3 (SUBJ-V-OBJ) RELATIONS
finding for a pair of tokens that have subj and obj relation respectively modifying a head
verb
a verb do occur was found. looking for its subject and object
a subject Bird flu and object invisible was found modifying the head verb do occur.
constructing a triple relation.
      Bird flu                                do occur                                invisible

TYPE 4 (PREPOSITIONAL MODIFIER) RELATIONS
finding for a preposition and its direct object that modifies either a noun or a verb
found the preposition in and the N invisible it modifies. the direct object of the
preposition is which country. constructing a triple relation.
      invisible                                in                                which country

DISCOURSE INTEGRATION FOR query
      POSSESSIVE RELATIONS
      APPOSITIVE RELATIONS
      SUBJ-B-OBJ RELATIONS
      Bird flu (0.2)                                do occur (0.4)                                invisible (0.0)
INPUT PREPOSITIONAL PHRASES
      invisible (0.E2)                                in (0.5)                                which country (0.7)
      ENTITY OBJECTS
      0.0                                entity(desc(X,invisible))
      0.2                                disease(disease_name(X,Bird flu))
      0.7                                variable(var_desc(X,which country))

INSTANTIATION OF ENTITY CLASSES AND THEIR ATTRIBUTES
instantiating a new entity class entity(X)
a new object entity(1519e762376c94457b985c9e86edf1ab) is obtained with the following
attributes:
      desc(1519e762376c94457b985c9e86edf1ab,invisible)
instantiating a new entity class disease(X)
a new object disease(43e7efee7c58619c29f1a2b86f857ae3) is obtained with the following
attributes:
      disease_name(43e7efee7c58619c29f1a2b86f857ae3,Bird flu)
instantiating a new entity class variable(X)
a new object variable(cf97ecfa503ad868c05bc035808b04cb) is obtained with the following
attributes:
      var_desc(cf97ecfa503ad868c05bc035808b04cb,which country)

TRIGGERS FOR EVENT CLASSES
attempt triggering using nouns (ONLY FOR QUERIES)
using the noun invisible(0.0)
using the noun Bird flu(0.2)
      the noun Bird flu(0.2) was triggered by the trigger word flu[outbreaks]
      the pattern for noun Bird flu(0.2) is no pattern
using the noun which country(0.7)
attempt triggering using verbs
using the verb do occur(0.4)
      the verb do occur(0.4) was triggered by the trigger word occur[outbreaks]
      the pattern for verb do occur(0.4) is {DISEASE}RELATION{LOCATION}
attempt triggering using prepositions
using the preposition in(0.5)
      the preposition in(0.5) was triggered by the trigger word in[outbreaks]
      the pattern for preposition in(0.5) is {VARIABLE|DISEASE}RELATION{LOCATION}

RESOLVING ANAPHORA

```

INSTANTIATION OF EVENT CLASSES AND THEIR ATTRIBUTES

instantiating a new event class **outbreaks** to obtain  
**outbreaks(1b1cc7f086b3f074da452bc3129981eb2)**

instantiating a new event class **outbreaks** to obtain  
**outbreaks(1b1cc7f086b3f074da452bc3129981eb2)**

instantiating a new event class **outbreaks** to obtain  
**outbreaks(1b1cc7f086b3f074da452bc3129981eb2)**

filling the attributes of event objects

relation **Bird flu(0.2)[outbreaks]** has the accompanying instances: this is a noun-triggered event. thus, no associated attributes, patterns and maps.

dynamically creating map

finding other trigger words of category **outbreaks** that has in\_pattern

relation **do occur(0.4)[outbreaks]** has the accompanying instances: **{Bird flu(0.2)[disease]}do occur{invisible(0.0)[entity]}**

the instances to the relation **do occur** FULFILLED the pattern **{disease}RELATION{location}** and the map **{disease\_name}RELATION{the\_location}** is used to fill the objects to the attributes of event object

disease in the first pattern matches the object disease(0.2)  
so, the object disease(43e7efee7c58619c29f1a2b86f857ae3) will be used to fill attribute **disease\_name**

location in the second pattern matches the object entity(0.0)  
so, the object entity(1519e762376c94457b985c9e86edf1ab) will be used to fill attribute **the\_location**

attributes-value pair collected so far for the event object  
**outbreaks(1b1cc7f086b3f074da452bc3129981eb2)**

=> **disease\_name(1b1cc7f086b3f074da452bc3129981eb2, 43e7efee7c58619c29f1a2b86f857ae3)**

    Ø **the\_location(1b1cc7f086b3f074da452bc3129981eb2,**

    Ø **1519e762376c94457b985c9e86edf1ab)**

relation **in(0.5)[outbreaks]** has the accompanying instances: **{invisible(0.E2)[variable]}in{which country(0.7)[variable]}**

the instances to the relation **in** DOES NOT fulfill the pattern **[variable|disease]RELATION[location]**

checking for inherited attributes among event objects

removing parent event objects

EVENTS AND THE RELATED ENTITY OBJECT OBTAINED FROM CURRENT SENTENCES

**outbreaks(1b1cc7f086b3f074da452bc3129981eb2)**

**disease\_name(1b1cc7f086b3f074da452bc3129981eb2,**

**43e7efee7c58619c29f1a2b86f857ae3)**

**disease(43e7efee7c58619c29f1a2b86f857ae3)**

**disease\_name(43e7efee7c58619c29f1a2b86f857ae3,**

**Bird flu)**

**outbreaks(1b1cc7f086b3f074da452bc3129981eb2)**

**the\_location(1b1cc7f086b3f074da452bc3129981eb2,**

**1519e762376c94457b985c9e86edf1ab)**

**entity(1519e762376c94457b985c9e86edf1ab)**

**desc(1519e762376c94457b985c9e86edf1ab,invisible)**

entity object attributes      **2 attributes**

entity objects                **2 objects**

event object attributes      **2 attributes**

event objects                 **2 objects**

OUTPUT OF DISCOURSE INTEGRATION: QUERY NETWORK

'1b1cc7f086b3f074da452bc3129981eb2','is','outbreaks'

'1b1cc7f086b3f074da452bc3129981eb2','disease\_name', '43e7efee7c58619c29f1a2b86f857ae3'

'43e7efee7c58619c29f1a2b86f857ae3','is','disease'

'43e7efee7c58619c29f1a2b86f857ae3','disease\_name','Bird flu'

'1b1cc7f086b3f074da452bc3129981eb2','is','outbreaks'

'1b1cc7f086b3f074da452bc3129981eb2','the\_location', '1519e762376c94457b985c9e86edf1ab'

'1519e762376c94457b985c9e86edf1ab','is','entity'

'1519e762376c94457b985c9e86edf1ab','desc','invisible'

NETWORK-BASED ADVANCED REASONING

question: Bird flu did occur in which countries

NETWORK-TO-PATH REDUCTION

from the query network, the set of path sequences  $Q$  from the leaf node to the root node is obtained:

path sequence  $q_i$ : Bird flu, disease\_name, 43e7efee7c58619c29f1a2b86f857ae3, disease\_name, 1blcc7f086b3f074da452bc3129981eb2, is, outbreaks

from the query network, the set of path sequence  $A$  from the leaf node **invisible** to the root node is obtained:

path sequence  $a_i$ : invisible, desc, 1519e762376c94457b985c9e86edflab, the\_location, 1blcc7f086b3f074da452bc3129981eb2, is, outbreaks

SELECTIVE PATH MATCHING & RELAXATION OF EVENT CONSTRAINT

using the semantic network, we try to verify that each path sequence in  $Q$  conditionally exists in  $S$

finding the path  $q_i$  Bird flu, disease\_name, 43e7efee7c58619c29f1a2b86f857ae3, disease\_name, 1blcc7f086b3f074da452bc3129981eb2, is, outbreaks from semantic network.  
The path is is **UNIQUE**

MATCHED the path bird flu, disease\_name, 43e7efee7c58619c29f1a2b86f857ae3, disease\_name, 6a8219886c22af2e6f91a0253fe70d1e2, is, outbreaks

MATCHED the path bird flu, disease\_name, 43e7efee7c58619c29f1a2b86f857ae3, disease\_name, 83dab367c7387e24b731198a990998321, is, outbreaks

MATCHED the path bird flu, disease\_name, 43e7efee7c58619c29f1a2b86f857ae3, disease\_name, a567fc0ef49e0d47316c0fc6de23de765, is, outbreaks

MATCHED the path bird flu, disease\_name, 43e7efee7c58619c29f1a2b86f857ae3, disease\_name, f9ca176513eb7ed1e6a73c780aa981863, is, outbreaks

potential **UNIQUE** event f9ca176513eb7ed1e6a73c780aa981863 from semantic network that appears in at least one path sequence  $q_i$

using the semantic network, proceed to find the value for **invisible** leaf node in set  $A$  of master event f9ca176513eb7ed1e6a73c780aa981863

the event f9ca176513eb7ed1e6a73c780aa981863 appears in 1 out of the 1 paths in  $Q$  and it is a **UNIQUE**.

finding the path invisible, desc, 1519e762376c94457b985c9e86edflab, the\_location, 1blcc7f086b3f074da452bc3129981eb2, is, outbreaks

MATCHED the path Malaysia, country(2), 756e795131d1efac6307a117aa7ec8ec, the\_location, f9ca176513eb7ed1e6a73c780aa981863, is, outbreaks

potential **UNIQUE** event 83dab367c7387e24b731198a990998321 from semantic network that appears in at least one path sequence  $q_i$

using the semantic network, proceed to find the value for **invisible** leaf node in set  $A$  of master event 83dab367c7387e24b731198a990998321

the event 83dab367c7387e24b731198a990998321 appears in 1 out of the 1 paths in  $Q$  and it is a **UNIQUE**.

finding the path invisible, desc, 1519e762376c94457b985c9e86edflab, the\_location, 1blcc7f086b3f074da452bc3129981eb2, is, outbreaks

MATCHED the path Germany, country(2), 2125c467080e7396e8e66d5dafc44196, the\_location, 83dab367c7387e24b731198a990998321, is, outbreaks

potential **UNIQUE** event a567fc0ef49e0d47316c0fc6de23de765 from semantic network that appears in at least one path sequence  $q_i$

using the semantic network, proceed to find the value for **invisible** leaf node in set  $A$  of master event a567fc0ef49e0d47316c0fc6de23de765

the event a567fc0ef49e0d47316c0fc6de23de765 appears in 1 out of the 1 paths in  $Q$  and it is a **UNIQUE**.

finding the path invisible, desc, 1519e762376c94457b985c9e86edflab, the\_location, 1blcc7f086b3f074da452bc3129981eb2, is, outbreaks

MATCHED the path Thailand, country(2), 2d12b535baf0916403a5d98dfe23fb6e, the\_location, a567fc0ef49e0d47316c0fc6de23de765, is, outbreaks

potential **UNIQUE** event 6a8219886c22af2e6f91a0253fe70d1e2 from semantic network that appears in at least one path sequence  $q_i$

using the semantic network, proceed to find the value for **invisible** leaf node in set  $A$  of master event 6a8219886c22af2e6f91a0253fe70d1e2

the event 6a8219886c22af2e6f91a0253fe70d1e2 appears in 1 out of the 1 paths in  $Q$  and it is a **UNIQUE**.

finding the path invisible, desc, 1519e762376c94457b985c9e86edflab, the\_location, 1blcc7f086b3f074da452bc3129981eb2, is, outbreaks

MATCHED the path Vietnam, country(2), b394edf1054b38ccee98af4663785a99, the\_location, 6a8219886c22af2e6f91a0253fe70d1e2, is, outbreaks

4 answer(s) were discovered: Malaysia, German, Vietnam, Thailand

**Note:**

The query results<sup>109</sup> found four answers (Malaysia, German, Vietnam and Thailand) from four differences trustworthiness websites extracted using AKEA as follows:

1. German - [www.pandemicflu.gov/general/workshopmorning.html](http://www.pandemicflu.gov/general/workshopmorning.html)  
MATCHED the path Germany, country(2), 2125c467080e7396e8e66d5dafc44196, **the location,** 83dab367c7387e24b731198a990998321, **is, outbreaks**
2. Malaysia - [www.whitehouse.gov/news/releases/2005/11/20051101-1.html](http://www.whitehouse.gov/news/releases/2005/11/20051101-1.html)  
MATCHED the path Malaysia, country(2), 756e795131d1efac6307a117aa7ec8ec, **the location,** f9ca176513eb7ed1e6a73c780aa981863, **is, outbreaks**
3. Vietnam - [www.who.int/csr/disease/avian\\_influenza/avian\\_faqs/en/index.html](http://www.who.int/csr/disease/avian_influenza/avian_faqs/en/index.html)  
MATCHED the path Vietnam, country(2), b394edf1054b38ccee98af4663785a99, **the location,** 6a8219886c22af2e6f91a0253fe70d1e2, **is, outbreaks**
4. Thailand [www2a.cdc.gov/HAN/ArchiveSys/ViewMsgV.asp?AlertNum=00221](http://www2a.cdc.gov/HAN/ArchiveSys/ViewMsgV.asp?AlertNum=00221)  
MATCHED the path Thailand, country(2), 2d12b535baf0916403a5d98dfe23fb6e, **the location,** a567fc0ef49e0d47316c0fc6de23de765, **is, outbreaks**

---

<sup>109</sup> Execution time for the query is **2.03 seconds** on Dell Precision PWS380 Server 3GH with 1GB of memory

## APPENDIX C: Google PageRank™ Checksum Calculator<sup>110</sup> - Assigning Numerical Weightings to Hyperlinked Documents Indexed by a Google Search Engine

```
/* Google PageRank™ Checksum Calculator */

header("Content-Type: text/plain; charset=utf-8");
define('GOOGLE_MAGIC', 0xE6359A60); //Define the named constant "GOOGLE_MAGIC"

//unsigned shift right
function zeroFill($a, $b)
{
    $z = hexdec(80000000);
    if ($z & $a)
    {
        $a = ($a>>1);
        $a &= (~$z);
        $a |= 0x40000000;
        $a = ($a>>($b-1));
    }
    else
    {
        $a = ($a>>$b);
    }
    return $a;
}

function mix($a,$b,$c) { // Google Checksum calculation
    $a -= $b; $a -= $c; $a ^= (zeroFill($c,13));
    $b -= $c; $b -= $a; $b ^= ($a<<8);
    $c -= $a; $c -= $b; $c ^= (zeroFill($b,13));
    $a -= $b; $a -= $c; $a ^= (zeroFill($c,12));
    $b -= $c; $b -= $a; $b ^= ($a<<16);
    $c -= $a; $c -= $b; $c ^= (zeroFill($b,5));
    $a -= $b; $a -= $c; $a ^= (zeroFill($c,3));
    $b -= $c; $b -= $a; $b ^= ($a<<10);
    $c -= $a; $c -= $b; $c ^= (zeroFill($b,15));

    return array($a,$b,$c);
}

//Calculate the Google Checksum from given URL
function GoogleCH($url, $length=null, $init=GOOGLE_MAGIC) {
    if(is_null($length)) {
        $length = sizeof($url);
    }
    $a = $b = 0x9E3779B9;
    $c = $init;
    $k = 0;
    $len = $length;
    while($len >= 12) {
        $a += ($url[$k+0]+($url[$k+1]<<8)+($url[$k+2]<<16)+($url[$k+3]<<24));
        $b += ($url[$k+4]+($url[$k+5]<<8)+($url[$k+6]<<16)+($url[$k+7]<<24));
        $c += ($url[$k+8]+($url[$k+9]<<8)+($url[$k+10]<<16)+($url[$k+11]<<24));
        $mix = mix($a,$b,$c);
        $a = $mix[0]; $b = $mix[1]; $c = $mix[2];
        $k += 12;
        $len -= 12;
    }

    $c += $length;
}
```

<sup>110</sup> Google PageRank™ Checksum Calculator script is originally from <http://www.googlecommunity.com/scripts/google-pagerank.php> and was modified and hosted at <http://ainibot.org/pagerank>



